Industrial Statistics

INDUSTRIAL STATISTICS

STATISTICAL TECHNIQUE APPLIED TO PROBLEMS IN INDUSTRIAL RESEARCH AND QUALITY CONTROL

H. A. Freeman

Associate Professor of Statistics
Massachusetts Institute of Technology

NEW YORK
JOHN WILEY & SONS, INC.
LONDON: CHAPMAN & HALL, LIMITED

FIFTH PRINTING, NOVEMBER, 1949

PRINTED IN THE UNITED STATES OF AMERICA

PREFACE

This book gives examples of the uses of elementary statistical methods in the design and analysis of experiments carried out in industrial plants and scientific laboratories. It also deals with several of the statistical features of the problem of establishing a systematic program through which the quality of industrial output can be studied and controlled. There is a final chapter on some of the statistical aspects of the relationship of sampling to the risks incurred by producers and buyers.

Those parts of the chapters in large type are meant to be usable by themselves; and they are intended for students, experimenters, and production men who are short on mathematical training. The notes in smaller type are included for the benefit of those who wish to go a little beyond the literary exposition of methods. These notes consist of comments on methods, rather detailed derivations, and mere outlines or suggestions of derivations. Some of the most important topics have been noted in the latter fashion for it is, unfortunately, true that many statistical techniques — which have long served industrial statisticians well — require rather advanced mathematics for their complete derivation.

The manuscript of this book has for the past several years formed the basis of a one-semester course in industrial statistics, Economics 38, given at the Massachusetts Institute of Technology. Students in this course are not expected to have had previous training in statistics.

The intermediate steps of many of the examples are omitted, but final answers are given. These partially complete examples can be used in assignments to students.

Those who work on industrial problems are aware of the obstacles to entirely successful use of statistical methods in industry. In particular, the lack of complete equivalence between industrial reality and our mathematical models thereof and the many technical complexities of manufacture and research make it advisable that our results be taken as tentative. Only those who are thoroughly familiar with the industrial or experimental process at hand can obtain the full benefits of the simple statistical methods described in this book and in other works of this character. In numerous instances in this book, my knowledge of the technical processes underlying the data under discussion is slight

and consequently my conclusions may have dubious practical significance.

I am deeply indebted to two former colleagues, Mr. Harold Bellinson, now of the War Department, and Mr. L. C. Young, now with Westinghouse, and to Mr. Churchill Eisenhart, of the University of Wisconsin, for many suggestions. Mr. Young and Margaret Z. Freeman have kindly carried out many of the computations. I am also indebted to our department secretaries, Miss Ethel Downer and Miss Eleanor Prescott, for typing the manuscript. Acknowledgments to those who have kindly permitted me to use their tables and their data are made elsewhere in the book.

I shall be glad to receive criticism and suggestions from readers.

H. A. FREEMAN

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS
May, 1942

CONTENTS

Statistical procedure and experimental design are only two different aspects of the same whole and that whole is the logical requirements of the complete process of adding to natural knowledge by experimentation.

R. A. FISHER
The Design of Experiments

CHAPTER I

THE DIFFERENCE OF TWO MEANS

1.1 Object of Chapter. We wish to design an experiment so that from two samples of data we can answer two questions: (1) are the averages in the two larger sources of data, from which the samples were drawn, equal or unequal; and (2) if they are unequal, within what limits can the value of the inequality be established? It is the purpose of this chapter to discuss conditions which should be satisfied by such an experiment, to illustrate a method of arranging the experiment so that even with a small number of observations the precision of inferences will be high, and finally to describe the relevant techniques for analyzing the data.

The design and analysis of experiments involving more than two averages will be considered in the following chapters.

1.2 Examples. Experiments having the objective stated above are performed in many branches of science. To give a few examples: in medicine and biology studies have been made of the effect of a certain amount of thymophysin (as compared to none) on blood pressure; also the difference in the number of bacterial colonies per plate when counted in the afternoon and in the evening. From industry and agriculture we have the difference in the effects of indoor and outdoor storage on the breaking strength of wood, comparison of the ash content of coals taken from two mines, and the difference in the yield of a particular variety of wheat under two types of fertilizer treatment.

1.3 Uses of the results. From such experiments two kinds of information may be wanted. First, what factor or factors are responsible for any observed difference in sample averages; and second, as a result of the experiment, what action should be taken? We shall consider these questions separately.

The difference in sample averages may be accidental rather than real, for samples can have different averages and yet the larger sources of data (to be known as *populations*), from which these samples were drawn, may have the same averages. If, however, the difference in sample averages is shown to be real, this difference may always be attributed separately to the influence of one or more factors and/or

to the joint influence of two or more factors. If the experiment is designed so that the two samples are unlike only with respect to one factor, that factor is held responsible for any real difference in averages that may be found. But the samples may be unlike with respect to several factors. For example, if it is shown that outdoor storage adversely affects the bending strength of wood, the factors responsible for the adverse effect may be sunlight and/or rainfall. It is even possible that the adverse effect is chiefly due to the joint action of sunlight and rainfall, these factors separately having slight influence. This preliminary experiment must then be followed by a further set of similar experiments in each of which both samples are alike with respect to all but one of the suspected factors; in this way, the responsible factor or factors can finally be identified.

It is possible, however, to plan the original experiment so that it alone will yield all this information; examples will be discussed in detail in the second chapter.

Laboratory, factory, and field experiments do not by themselves provide sufficient information to determine economic policy. For example, an experiment shows that the bending strength of wood is impaired by outdoor storage. Users of wood may, however, be partly interested in another quality characteristic, such as hardness, which by a similar experiment can be shown to be unaffected by outdoor storage. Users will, therefore, be willing to pay only a fraction of the premium arising from the additional cost of indoor storage and the determination of that fraction clearly depends on facts not supplied by either experiment. If between two methods of manufacture or two types of product no real (statistical) difference is found, users will have no definite preference and producers will favor the method or product involving the lesser cost. When a real difference is found, and if that difference is *practically* significant, the resultant shift in the preference of users, as well as cost differences, will determine the effect on the market of the results of the experiment.

1.4 Problems facing the experimenter. We shall now consider a specific example, but the reader should be able to apply the discussion to any experiment involving a difference of two averages.

An experimenter wishes to determine whether or not the average amounts of corrosion of two types of wrought ferrous pipe coating are the same. The two types are open-hearth iron and puddled iron. He selects several specimens of each coating, buries them in the soil and, on later removing them, measures the corrosion of each specimen. If the amount of corrosion of open-hearth coating is designated by the variable X and that of puddled-iron coating by the variable Y,

his data on p specimens of the former and q specimens of the latter are as follows:

$$\begin{array}{cc} X_1 & Y_1 \\ X_2 & Y_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ X_p & \cdot \\ & Y_q \end{array}$$

In carrying out this experiment, he will have had to make several important decisions, and on them depends much of the reliability of his later inferences. First, should factors that can be held constant be allowed to vary? For example, should all specimens of pipe coating be of the same size, be buried in the same type of soil, at the same depth, covered with the same backfill, and be left in the soil the same period of time? Second, what account can be taken of uncontrollable factors, such as the weather after the burial of the specimens? Third, how are test specimens to be selected from the larger sources of supply, how many should be taken, and should there be a like or unlike number of each type of coating? Fourth, what index of corrosion shall be used? This is only a partial list but it covers the types of questions that must be answered in any experiment of this kind.

1.5 Desirability of control. Let us first compute the *arithmetic means* $\bar{X}$ and $\bar{Y}$ of the two samples. These averages are given by

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_p}{p}$$

$$\bar{Y} = \frac{Y_1 + Y_2 + \cdots + Y_q}{q}$$

and they can be regarded as estimates of the respective population means $\bar{X}'$ and $\bar{Y}'$. We will consider the magnitude of $\bar{X} - \bar{Y}$ in the light of the tentative hypothesis that the population means $\bar{X}'$ and $\bar{Y}'$ are equal. If $\bar{X} - \bar{Y}$ is not zero and if we can show that its departure from zero was not accidental, the hypothesis $\bar{X}' - \bar{Y}' = 0$ will be rejected.

Confidence that $\bar{X} - \bar{Y}$ is an accurate measure of $\bar{X}' - \bar{Y}'$ is in part dependent on the variability among observations in the populations from which the two samples are drawn. Assume that a sample can be drawn so that the unknown variability of the variates in the population

is proportional to the calculated variability of the variates in the sample. If, then, the sample variates $X_1, X_2, \cdots, X_p$ differ slightly from each other, $\bar{X}$ will be a relatively reliable measure of the population mean $\bar{X}'$ in the sense that further sampling from the population would not greatly affect $\bar{X}$. If $X_1, X_2, \cdots, X_p$ vary considerably, $\bar{X}$ is a less reliable estimate of $\bar{X}'$. This would be the case if, for example, open-hearth coatings were buried in several types of soils which differ in their corrosiveness or if some specimens were removed from the soil before others. The argument is similar for Y.

1.6 Pairing. In the example given, variability among $X_1, X_2, \cdots, X_p$ and among $Y_1, Y_2, \cdots, Y_q$ can be reduced to a minimum by using specimens all of the same size, burying them in the same soil and for the same length of time, etc. The precision of inference will be thereby improved, but the great disadvantages of this type of experiment lies in its reduced generality and in the practical difficulty of performing such an experiment at all. Complete control over all relevant factors — pipe size, kind of soil, and burial period — is practically impossible to achieve in ordinary experimenting.

The arrangement shown in the following table attains both objectives — practicality and precision. First, it makes possible the introduction into the experiment of variability in such important factors as type of soil and period of burial, thus making the experiment practically feasible and allowing it to simulate conditions of industrial life; and second, it excludes the influence of the variability of these factors on the precision of inferences relating to the arithmetic means.

Kind of soil, length of burial	Corrosion		
	Open-hearth iron coatings	Puddled-iron coatings	Difference
Clay, A years	X_1	Y_1	$d_1 = X_1 - Y_1$
Cinders, B years	X_2	Y_2	$d_2 = X_2 - Y_2$
.	.	.	.
.	.	.	.
.	.	.	.
Loam, C years	X_n	Y_n	$d_n = X_n - Y_n$
Mean	$\bar{X}$	$\bar{Y}$	$\bar{d} = \bar{X} - \bar{Y}$

Each of the quantities $d_1, d_2, \cdots, d_n$ is unaffected by differences among various soils and the various lengths of burials, for in each pairing both kinds of pipe are treated alike with respect to these factors. Hence the error of $\bar{d}$ tends to be small. At the same time, the experi-

ment manages to include the various soil types and lengths of burial occurring in the ordinary industrial use of these coatings. Even the uncontrollable factor, variable weather, is also introduced and its effect likewise excluded, for the two specimens in any pairing will likely be buried side by side. One must, however, recognize the fact that the results of this experiment are not as reliable for a *particular* combination of the influential factors — say cinders, B years burial and heavy rain after burial — as an experiment in which all $2n$ observations were devoted to that combination.

In a single experiment a very wide range in the nature of these combinations of influential factors is disadvantageous. Thus, in muck soil, the superiority of open-hearth coverings may be much greater than in any other soil, that is, the value of d will be relatively large in that pairing. The increased freedom allowed the experimenter by the inclusion of this kind of soil may be offset by the increased variability of the variates d_i. If this unusual reaction to muck soil is already familiar to the experimenter, then muck soil should not be included in the present experiment, for its inclusion is uninformative and the loss in precision is costly.

It is important to note that the estimate, from the paired results, of the true error of the *mean difference* $\bar{d}$ is based on the n variates d_i whereas in the case in which $\bar{d}$ was formed from unpaired observations, the estimate of error is based on the $2n$ variates X_i and Y_i. In each example we shall have to determine whether the increased precision of $\bar{d}$ resulting from the reduction of variability due to pairing is or is not offset by the loss of precision due to a 50 per cent reduction in the number of variates.

1.7 Randomization. The two objectives, precision and practicality, are achieved by pairing. The remaining objective is to avoid bias, and this can be achieved by randomization.

It may happen that certain influential factors cannot be handled by pairing. In such a case, the influences of these factors cannot be eliminated, but they can be distributed so that our comparison of $\bar{X}$ and $\bar{Y}$ is not vitiated by their presence. To illustrate the point, assume that in the present experiment the orientation of specimens in the soil might influence the amount of their corrosion. If, then, all open-hearth specimens are buried in the east side of each excavation and all puddled-iron specimens in the west side, any conclusion that, say, open-hearth coatings are better than puddled-iron coating is now assailable on the ground that the east side may have been a favorable position. This possibility can be precluded simply by assigning positions to the specimens of each pairing in random fashion, for example, by tossing a coin.

Randomization provides a completely objective technique of removing the possible systematic effects of uncontrolled factors, effects which if not randomized might vitiate the comparison of the means.

Size of pipe, i.e., exposed area, somewhat affects corrosion as measured by depth of pits. This factor has not been included in the pairing arrangement for the resultant need of drawing a fixed number of each size of pipe from the population would interfere with the simple sampling technique to be discussed in the next section. This factor should therefore be randomized. The same argument applies to any factor. If length of burial is not included among the controls in the pairing arrangement, random selection of burial periods will preclude the possibility that this factor will vitiate the results — something which might happen if longer burial periods were unfortunately associated with one kind of coating. It is disadvantageous to randomize a factor which could be controlled by pairing, for the effect is to increase the variability among the d_i and therefore the error against which $\bar{d}$ is judged.

1.8 Selection of specimens. The method of selecting specimens of each type of coating must be one which will not vitiate the experimenter's conclusions. For example, if, as a result of biased sampling, the open-hearth coatings used in the experiment are better on the average than those in their population while puddled-iron specimens are, on the average, poorer than those in their population, any inference of the nature of $\bar{d}'(= \bar{X}' - \bar{Y}')$ from the observed data will be vitiated. As a second example, if as a result of biased sampling, the open-hearth specimens in the sample are more uniform in amounts of corrosion than the specimens of their population, an incorrectly high precision will be placed on $\bar{X}$.

Such bias can be avoided by selection of the specimens for each sample in such a way that all specimens of the corresponding population have an equal opportunity of being drawn. Such random selection may be carried out in the following way: Assume there are 30,000 open-hearth specimens in the population and 40 are to be drawn. Assign numbers 1 to 30,000 to the specimens in the population. From any page of a table of random numbers (numbers composed of randomly selected digits) write down in order five-place numbers (omitting all numbers over 30,000) until the numbers of 40 specimens have been drawn. Similarly for puddled-iron specimens. Among such tables is one by Fisher and Yates (16) in which the digits were obtained from the 15th to the 19th digits of a set of 20-place logarithms. The direct approach would be to draw at random from a well-mixed bowl of 30,000 chips marked from 1 to 30,000, but the labor of marking is great.

The purpose of random selection is clear but random selection in practice may be difficult. For example, in dealing with fibers it is impossible to assign a number to each specimen in the population; furthermore precaution will have to be taken to avoid the tendency to draw the longer fibers. Chance may select a specimen from the center of a rug, or the specimen of pipe whose number is drawn may be at the bottom of a pile of thousands of specimens. These difficulties necessitate compromises but every effort should be made to remove subjective decision and its attendant biases from the method of selection.

1.9 Size of the experiment. The number of specimens to be used in an experiment is related to (*a*) the expected value of the mean difference, (*b*) the variability of the variates in the population, and (*c*) the confidence with which our conclusions are to be stated. If in two experiments factors (*a*) and (*b*) are the same, then the greater the desired degree of confidence, the larger must be the size of the samples. If (*a*) and (*c*) are the same, the greater the variability, the larger the size of the samples. If (*b*) and (*c*) are the same, then the greater the expected value of the mean difference, the smaller the size of the samples.

Another important influence on the size of the experiment results from the fact that the variability of the variates in the population, whether large or small, must be estimated from the samples. This estimate is subject to error, and this error is reduced by use of larger samples.

These general considerations do not enable an experimenter to decide whether he will need 10 or 50 specimens. Full information on factors (*a*) and (*b*) may be available only when the experiment is completed. If advance estimates of the magnitudes of (*a*) and (*b*) can be made, the proper value of the size of each sample, n, can be approximated by formulae to be developed presently.

1.10 Quality characteristics. Users of an industrial product are often interested in more than one of its qualities. For example, both hardness and tensile strength may be important. Two possibilities are open to the experimenter: (*a*) he may conduct separate experiments for each quality characteristic or, (*b*) if not more than one quality characteristic necessitates a destructive test, he can obtain data on all characteristics from one experiment. In the case of hardness and tensile strength (*b*) would apply, for the test for hardness is not destructive.

For any one quality characteristic several measures may be available. For example, corrosion may be measured by loss of weight or by depth of maximum pits. The experimenter should generally choose a measure which varies continuously in preference to one which can assume only a

few values. An experiment on corrosion using loss of weight or depth of pits (both of which vary continuously upwards from zero) yields more information than a like-sized experiment in which amount of corrosion is measured simply as high, medium, and low. Such a crude classification conceals information which a continuous measure reveals. We shall not discuss methods appropriate to this crude type of classification, although for a few industrial products it may be the only type available.

1.11 An experiment in detail. Thirty specimens, fifteen of each type of coating, are drawn at random from their respective populations. One specimen of each type of coating is included in each pair; each pair is buried in the same soil, in similar positions, at the same depth and for the same period of time. The various pipe sizes, ranging from 1 inch to 1½ inches, are randomized. The results follow:

Controls		Depth of maximum pits (expressed in thousandths of an inch)		
Kind of soil	Length of burial (years)	Open-hearth iron coatings	Puddled-iron coatings	Difference
Clay	4.5	73	51	+22
Clay	3.8	43	41	+ 2
Cinders	7.1	47	43	+ 4
Cinders	6.1	53	41	+12
Peat	2.0	58	47	+11
Tidal marsh	4.4	47	32	+15
Loam	5.5	52	24	+28
Clay	9.2	38	43	− 5
Clay	8.5	61	53	+ 8
Clay	8.0	56	52	+ 4
Loam	5.7	56	57	− 1
Clay	3.2	34	44	−10
Clay	4.2	55	57	− 2
Loam	6.6	65	40	+25
Alkali knoll	6.4	75	68	+ 7

1.12 General nature of the test of the hypothesis $\bar{d}' = 0$. The test of the hypothesis $\bar{d}' = 0$ proceeds as follows: First, considerable information regarding the distribution of d_i in the population is assumed to be at hand. Now assume that from this population of d_i a very large number k of random samples each of n specimens have been drawn and the mean of each sample computed. It will be found that means which depart considerably from the population mean ($\bar{d}' = 0$) occur infrequently whereas means near $\bar{d}' = 0$ occur frequently. The frequency

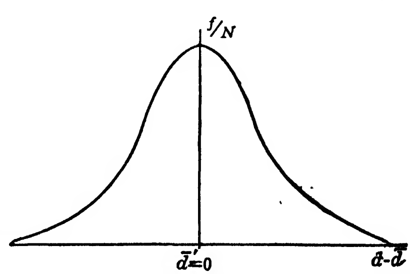
g and the fractional frequency or probability g/k of samples whose means depart from the population mean by a given amount is thus experimentally determinable. We may then determine by actual count the probability P of a departure from the population mean as large or larger than that actually observed. As a matter of fact, the distribution of sample means is mathematically determinable, so there is no need for the laborious experimental approach to the determination of P. If P is large, the observed difference $|\bar{d} - \bar{d}'|$ is attributed to the vagaries of sampling; if P is small, the difference $|\bar{d} - \bar{d}'|$ is taken to be real and the hypothesis $\bar{d}' = 0$ is rejected; the two materials under investigation are said to be significantly different in their means.

1.13 Normality of the population. One of the facts assumed to be known of the population is that the frequencies f_i of values of d_i in the population are normally distributed; that is,

$$f = \frac{N}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{d-\bar{d}'}{\sigma}\right)^2} \tag{1}$$

where N is the total frequency (total number of observations) in the population (N can be assumed to be infinite) and σ is the *standard deviation* of d, the nature of which will be discussed presently. Just as the equation $y = \alpha x$ represents a straight line with slope depending on the value of the parameter α, so [1] represents a normal distribution of frequencies with exact shape and position depending on values of the parameters $\bar{d}'$ and σ.

Three simple properties of [1] may be noted here. The squared exponent of e shows that the frequencies of $+(d - \bar{d}')$ and $-(d - \bar{d}')$ are equal for any d, that is, the distribution is symmetrical around $d = \bar{d}'$. The maximum frequency occurs when the exponent of e is zero, which is at $d = \bar{d}'$. Finally f approaches zero as $\pm(d - \bar{d}')$ becomes large. If the probability f/N is plotted against the deviation $d - \bar{d}'$, we have the following curve, for fixed values of the parameters $\bar{d}'$ and σ.



The technique of testing the hypothesis that the population is normal is similar in general nature to the test of the hypothesis $\bar{d}' = 0$ and to practically all other tests that will be made in this book. From the data of the sample we compute one or more constants whose values are known for a perfectly normal population. Then allowance is made for

the fact that even from a perfectly normal population a random sample having non-normal characteristics may by chance be drawn, particularly if the size of the sample is small. If, for the particular value of n used, the departures of the sample constants from the known normal values are greater than can be so allowed by chance the hypothesis that the sample came from a normal population is rejected.

Two such constants* are

$$\sqrt{\beta_1} = \frac{\text{Third moment of the population about its mean}}{(\text{Second moment of the population about its mean})^{3/2}} = \frac{\dfrac{\sum^{N}(d - \bar{d}')^3}{N}}{\left[\dfrac{\sum^{N}(d - \bar{d}')^2}{N}\right]^{3/2}}$$

and

$$\alpha = \frac{\text{Mean deviation of the population}}{(\text{Second moment of the population about its mean})^{1/2}} = \frac{\dfrac{\sum^{N}|d - \bar{d}'|}{N}}{\left[\dfrac{\sum^{N}(d - \bar{d}')^2}{N}\right]^{1/2}}$$

For a normal distribution $\sqrt{\beta_1} = 0$ and $\alpha = \sqrt{2/\pi}$. The former is obvious because a normal distribution is symmetrical around its mean; hence any odd moment about the mean will be zero. Assume that from a normal population a very large number of samples, each of size n, are drawn and for each sample two constants $\sqrt{b_1}$ and a are computed, where

$$\sqrt{b_1} = \frac{\text{Third moment of the sample about its mean}\dagger}{(\text{Second moment of the sample about its mean})^{3/2}} = \frac{\dfrac{\sum^{n}(d - \bar{d})^3}{n}}{\left[\dfrac{\sum^{n}(d - \bar{d})^2}{n}\right]^{3/2}}$$

and

$$a = \frac{\text{Mean deviation of the sample}}{(\text{Second moment of the sample about its mean})^{1/2}} = \frac{\dfrac{\sum^{n}|d - \bar{d}|}{n}}{\left[\dfrac{\sum^{n}(d - \bar{d})^2}{n}\right]^{1/2}}$$

and

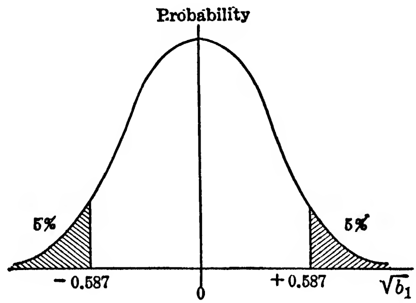
If the resulting distribution of the frequencies of values of $\sqrt{b_1}$ is examined, it will be found that values near zero occur most frequently.

* These parameters should be defined here and elsewhere in this chapter in terms of integrals but the summations used here should cause no difficulty.

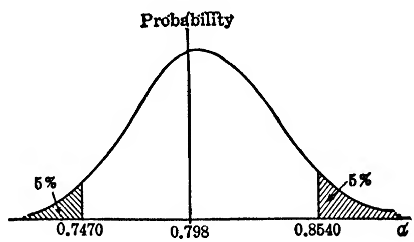
† Properly, third moment of the elements of the sample about their mean, etc.

The form of this distribution has been successfully approximated and Table I shows values of $\sqrt{b_1}$, for given n, beyond which 5 per cent and 1 per cent of all values of $\sqrt{b_1}$ of random samples from a normal population are found. Similar information on a is shown in Table III.

It will be noted that 5 per cent and 1 per cent of the frequencies may be interpreted as 5 per cent and 1 per cent of the area under the frequency curve. Thus, for $n = 40$ the graph of $\sqrt{b_1}$ illustrates the situation.



Similar arguments hold for a, the distribution of which is not symmetrical about $\sqrt{2/\pi}$ ($= 0.798$). Thus for $n = 41$, the graph of a illustrates the situation.

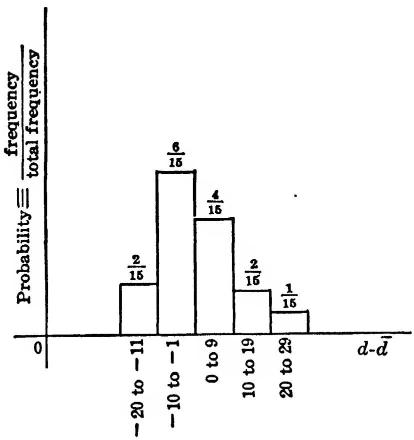


In the present example on the corrosion of pipe coatings we find

$$\sqrt{b_1} = 0.330$$
$$a = 0.814$$

If fewer than say 1 per cent or 2 per cent of random samples of size $n = 15$ yield values departing by as much as or more than 0.330 and 0.016 from the expected values 0 and 0.798, the sample at hand cannot be considered to have been drawn from a normal population. From Table I, 1 per cent of all random samples have values of $\sqrt{b_1}$ exceeding 1.061 (for sample size 25, the first entry in the table). Now, the spread

of the distribution of the $\sqrt{b_1}$ is less for large than for small n. Hence, more than 1 per cent of all random samples of size 15 have $\sqrt{b_1} > 0.330$, and the hypothesis of normality is not refuted.



From Table III, the 1 per cent levels of a are approximately 0.92 and 0.68. Our value $a = 0.814$ is within this range. Hence the hypothesis of normality is not refuted by this second (and independent) test.

For small samples these tests are sensitive only to large departures from normality. The diagram shown below of the sample data appears somewhat non-normal in skewness but the $\sqrt{b_1}$ test, which is a test of skewness, did not offer support.

1.14 Variance of the population. It has been noted that the reliability of $\bar{d}$ depends in part on the variability in corrosion of the specimens in the population, so a knowledge of the amount of this variability is necessary. One measure which would seem reasonable is

$$\frac{\sum^{N}(d - \bar{d}')}{N}$$

but this is not useful, for its value is always zero.

$$\sum^{N}(d - \bar{d}') = \sum^{N} d - N\bar{d}' = N\bar{d}' - N\bar{d}' = 0$$

A second possibility is the average of the sum of the absolute values of deviations of observations about their mean, i.e., the *mean deviation*

$$\frac{\sum^{N}| d - \bar{d}' |}{N}$$

This is algebraically an inconvenient measure and it does not fit well into the general body of statistical theory. The best measure of variability is the standard deviation σ, which has already been introduced as a parameter of the normal distribution.

$$\sigma = \sqrt{\frac{\sum^{N}(d - \bar{d}')^2}{N}}$$

We shall work with the *variance* σ^2. The true value of the population variance σ^2 is unknown, but it can be estimated from the sample data. If the sample at hand is large, a good estimate ($\hat{\sigma}^2$) of σ^2 is the sample variance s^2, which is given by

$$[2] \qquad \frac{\sum^{n} (d - \bar{d})^2}{n}$$

If the sample is small, we shall later show in part that the appropriate estimate $\hat{\sigma}^2$ is given by

$$[3] \qquad \frac{\sum^{n} (d - \bar{d})^2}{n - 1}$$

the divisor representing the number of independent values of d (degrees of freedom). Thus from n values of d one constant $\bar{d}$ has been calculated; hence, given this value of $\bar{d}$, only $n - 1$ values of d are unfixed, or independent. [3] is always better than [2] but for samples with $n > 30$ the difference may be neglected. We write

$$\sum (d - \bar{d})^2 = \sum d^2 - 2\sum d\bar{d} + \sum \bar{d}^2$$

the last two terms of the above can be combined and we have

$$\sum (d - \bar{d})^2 = \sum d^2 - n\bar{d}^2 = \sum d^2 - \frac{(\sum d)^2}{n}$$

where n is the number of observations. The last form is the most convenient for the purposes of calculation.

In the present example

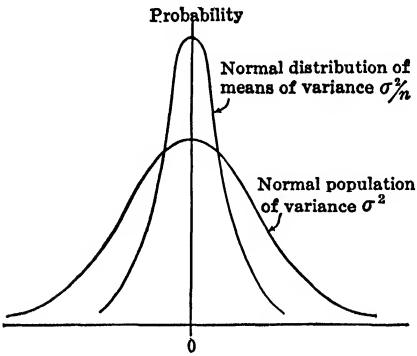
$$\hat{\sigma}^2 = 121.571$$

1.15 The *u* test. The appropriate tests of the difference of two means may now be described in greater detail. Assume that we are given a normal population of *known* variance σ^2 and mean $\bar{d}'$; what is the distribution of the means of random samples each of n observations?

First, consider a population defined only by $\bar{d}' = 0$. If from such a population a very large number k of samples each of size n are drawn, and the fractional frequencies or probabilities g_i/k of their means $\bar{d}_i$ are calculated, a distribution of the frequency of the various means can be plotted. It will be apparent to the reader that (1) the maximum frequency of this distribution occurs at $\bar{d} = \bar{d}'(= 0)$, (2) the distribution of sample means has smaller variance than the parent population, (3) the variance of the distribution of means is smaller for large than for small n; and (4) if the population is symmetrical about $\bar{d}'$ (as is the

normal distribution) the distribution of sample means will be symmetrical about $\bar{d} = \bar{d}'$.

In support of (2) and (3) it will later be proved that if the variance of the population is σ^2, the variance of the distribution of means is σ^2/n. In connection with (4) it will be shown that if the population is normal the distribution of the sample means will also be normal.



If the fractional frequency or probability of random samples with means greater than the mean of our sample $\bar{d}$ (and less than $-\bar{d}$) is low, say less than 5 per cent, the hypothesis that our sample was a random sample from a normal population of mean $\bar{d}' = 0$ and variance σ^2 is hardly tenable. If this is found and if (1) our sample is random and (2) the variance of the population is σ^2 and (3) the population is normal, it follows that $\bar{d}' \neq 0$. The means of the two materials are significantly different.

This test of differences of means, which will be called the u test, is thus based on the following theorem: Given a normal population of mean $\bar{d}'$ and variance σ^2, the means of random samples each of n observations will be distributed normally with mean $\bar{d}'$ and variance σ^2/n. This may also be expressed as follows: given a normal population of mean $\bar{d}'$ and variance σ^2, the statistic $u = \dfrac{\bar{d}}{\sigma_{\bar{d}}}\left(= \dfrac{\bar{d}}{\sigma/\sqrt{n}}\right)$ is normally distributed with mean $\bar{d}'$ and variance unity.

1.16 The *t* test. In our case the variance of the normal population is unknown; it must be estimated from a small sample, and the u test must be modified. The statistic

$$t = \frac{\bar{d}}{\hat{\sigma}/\sqrt{n}}$$

will be distributed symmetrically with mean $\bar{d}'$, but the distribution will be somewhat more peaked and will, in general, have a wider range than the normal distribution, with its shape depending on the number of independent observations (called degrees of freedom) from which the estimate $\hat{\sigma}^2$ is calculated. The test of significance is called the t test, and the peaked distribution is known as " Student's " distribution.

In our example the population of differences is normal; also $\bar{d}' = 0$, by hypothesis, and $\hat{\sigma}^2 = 121.571$; the variance $\sigma_{\bar{d}}^2$ of the distribution of means is $\hat{\sigma}^2/n = 8.105$ and the standard deviation (standard error) of the distribution of means is $\sqrt{8.105} = 2.847$. The difference $\bar{d} - \bar{d}'$ is 0.008 inch. We express this difference in terms of a measure of its error such as $\sigma_{\bar{d}}$; this division of one linear function $(\bar{d} - \bar{d}')$ by another, $\sqrt{\dfrac{\sum(d - \bar{d})^2}{n(n-1)}}$, eliminates the effect of the units of the difference and permits the use of a single set of tables for all problems. The difference of 0.008 inch is 8/2.847 or 2.81 standard error units.

The deviation is 2.81 standard error units, and the estimate of the variance is based on 14 degrees of freedom. From Table V the probability of exceeding by chance a deviation of 2.81 standard error units is only 0.015, approximately. Thus the two kinds of pipe are significantly different in their rates of corrosion. If the two types differ only in one characteristic, say method of manufacture or inclusion or exclusion of slag, this factor can be held responsible for the difference in quality.

1.17 Analysis of unpaired variates. If kind of soil and length of burial do not affect corrosion, it is disadvantageous to consider specimens to be paired with respect to these factors, for the variates d_i will be no less variable than X_i and Y_i, and there are only n variates d_i in place of $2n$ variates X_i and Y_i. If kind of soil and length of burial affect corrosion, pairing will likely be advantageous. To determine the gain or loss resulting from pairing, the variates, considered as unpaired, must be analyzed.

Given a normal population of mean $\bar{X}' - \bar{Y}' = 0$ and of variance σ^2; assume that from this population two random samples are drawn, of size n_X and n_Y, and that the difference of their means is $\bar{X} - \bar{Y}$. If a large number k of such dual drawings are made, the resulting k values of $\bar{X} - \bar{Y}$ may be grouped into a frequency distribution. It should be apparent that (1) this distribution of means will center about $\bar{X}' - \bar{Y}' = 0$; (2) the most frequently occurring value of $\bar{X} - \bar{Y}$ will be zero; (3) the distribution will be symmetrical about $\bar{X}' - \bar{Y}' = 0$; (4) it will probably have smaller variance than the population, and (5) its variance is small when n_X and n_Y are large. It will later be proved that this distribution of the difference of means is normal with variance

$$\sigma^2\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)$$

One important difference between the analysis of paired and unpaired variates lies in the estimate of the population variance. In the case of paired variates, we have a single sample and the estimate $\hat{\sigma}^2$ is given by

$$\frac{\sum(d - \bar{d})^2}{n - 1} \equiv \frac{\sum[(X - \bar{X}) - (Y - \bar{Y})]^2}{n - 1}$$

With unpaired variates the estimate $\hat{\sigma}^2$ will be shown to be

$$\frac{\sum^{n_X}(X - \bar{X})^2 + \sum^{n_Y}(Y - \bar{Y})^2}{n_X + n_Y - 2}$$

Note that the estimate $\hat{\sigma}^2$ from a single sample of differences is based on $n - 1$ independent differences whereas if the original variates are not, or are considered not to be, paired, the estimate is based on $n_X + n_Y - 2$ independent variates.

We have, for the unpaired variates

$$\bar{d}' = 0$$
$$\bar{d} = 8$$
$$n = 15$$
$$\hat{\sigma}^2 = 125.029$$

$$\hat{\sigma}^2\left(\frac{1}{n} + \frac{1}{n}\right) = 16.671$$

and the standard error of the difference of means is

$$\hat{\sigma}\sqrt{\frac{1}{n} + \frac{1}{n}} = 4.08$$

The deviation in standard error units is

$$\frac{8}{4.08} = 1.96$$

which for 28 degrees of freedom is not significant, for P is greater than 0.05, whereas in the analysis of the paired variates, the difference was significant. In this example the gain in sensitivity from pairing outweighed the loss of half of the degrees of freedom, and the testimony of the paired variates may be accepted. This gain in sensitivity resulted from the exclusion, by pairing, of the effects of factors which affected both samples.

1.18 Equality of the variances. In testing the significance of the difference of the means of unpaired variates; the statistic t was computed, where

$$t = \frac{\bar{d}}{\sqrt{\dfrac{\sum(X - \bar{X})^2 + \sum(Y - \bar{Y})^2}{n_X + n_Y - 2}\left(\dfrac{1}{n_X} + \dfrac{1}{n_Y}\right)}}$$

This might have been written

$$t = \frac{\bar{d}}{\sqrt{\dfrac{n_X s_X^2 + n_Y s_Y^2}{n_X + n_Y - 2}\left(\dfrac{1}{n_X} + \dfrac{1}{n_Y}\right)}}$$

where s_X^2 and s_Y^2 are the sample variances. The test of the means of unpaired variates results in the acceptance or rejection of the hypothesis that the two normal populations have the same mean $\bar{X}' = \bar{Y}'$ and the same variance $\sigma_X^2 = \sigma_Y^2$. If $\sigma_X^2 \neq \sigma_Y^2$ any inference regarding the validity of the hypothesis $\bar{X}' = \bar{Y}'$ is open to question, for a large value of t may reflect differences in variances rather than differences in means. To test the hypothesis $\sigma_X^2 = \sigma_Y^2$ we compute from the two samples the value of the statistic L_1 which for k samples is given by

$$L_1 = \left(\frac{s_1^2 s_2^2 \cdots s_k^2}{s_a^2 s_a^2 \cdots s_a^2}\right)^{1/k}$$

where

$$s_a^2 = \frac{1}{k}\sum_{i=1}^{k} s_i^2$$

If the variances are identical, $s_1^2 = s_2^2 = \cdots = s_a^2$, then $L_1 = 1$. The distribution of L_1 for random samples from a normal population has been approximated and Table X shows values of L_1, for samples of various sizes, beyond which 5 per cent and 1 per cent of all values of L_1 lie. Note that L_1 always lies between 1 and 0. We have

$$s_1^2 = 125.09$$
$$s_2^2 = 108.29$$
$$s_a^2 = 116.69$$

from which

$$L_1 = 0.997$$

This is far above the 5 per cent level of L_1 shown in Table X (the 5 per cent level of L_1 is 0.8673); accordingly the variances of the two normal populations from which these samples were drawn are not significantly different.

1.19 Other tests of significance of the difference of two means. If the population is definitely not normal, it is necessary to use a test not assuming normality. One such test has been given by Wald and Wolfowitz (45). If the L_1 test indicates that the variances differ significantly, a test of the hypothesis $\bar{X}' = \bar{Y}'$ has been proposed by Fisher and Behrens (Sukhatme, 40, for examples, tables). Often in experimental work, both normality and equality of variances will be found or can be assumed, and in such instances Student's t test, which incorporates normality and equal variances into its hypothesis, should be used; the t test (or for large samples, the u test) under these conditions will be more sensitive than a test which is designed to be valid for more general conditions.

1.20 Further examples. Beckwith (2) gives the following data for tuft bind tests on each of two rugs. The values are unpaired; only one test of significance is available.

Rug No. 1	Rug No. 2
10.0	10.5
10.5	9.5
9.5	8.5
18.5	9.0
14.0	8.5
14.0	12.0
12.0	8.0
9.5	10.5
12.5	7.0
10.0	10.5

Are the population means significantly different? The test already used may be summarized as follows: If a large number of pairs of small samples of size n_X and n_Y respectively are drawn at random from a normal population of mean $\bar{d}' = \bar{X}' - \bar{Y}'$ and variance σ^2, the quantity $\bar{d}/\sigma_{\bar{d}}$ is distributed as Student's t with $n_X + n_Y - 2$ degrees of freedom, where

$$\bar{d} = \bar{X} - \bar{Y}$$

$$\sigma_{\bar{d}}^2 = \sigma^2\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)$$

and the best estimate $\hat{\sigma}^2$ of the unknown variance σ^2 is

$$\frac{\sum(X - \bar{X})^2 + \sum(Y - \bar{Y})^2}{n_X + n_Y - 2}$$

We have

$$\hat{\sigma}^2 = 5.174$$
$$\sigma = 2.275$$

$$t = \frac{\bar{d}}{\sigma_{\bar{d}}} = \frac{\bar{d}}{\hat{\sigma}\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} = \frac{2.65}{2.275\sqrt{\frac{2}{10}}} = 2.605$$

For $t = 2.605$ with 18 degrees of freedom we find from Table V that $P < 0.02$. The means are significantly different.

The following data showing the results of field tests on the corrosion of non-bituminous pipe coatings for underground use have been given by Logan and Ewing (25).

Soil type	Lead coated steel pipe	Bare steel pipe
A	27.3	41.4
B	18.4	18.9
C	11.9	21.7
D	28.7	9.8
E	11.3	16.8
F	14.8	9.0
G	20.8	19.3
H	21.6	11.1
I	17.9	32.1
J	7.8	7.4
K	18.6	68.3
L	14.7	20.7
M	19.0	34.4
N	65.3	76.2

Do these two types of pipe differ significantly in their resistance to corrosion?

Analysis of the unpaired variates yields

$$t = -0.931$$

which for 26 degrees of freedom is not significant.

In this example there is some evidence that the data in any one row are not independent. Type of soil is probably responsible for any such lack of independence; for example, soil N appears to be highly corrosive to *both* types of coatings whereas soil J has slight effect regardless of covering; this "positive" correlation is, however, not in evidence in all pairs. Analysis of the paired variates gives the results shown on page 20.

Soil type	Difference in penetration
A	−14.1
B	− 0.5
C	− 9.8
D	+18.9
E	− 5.5
F	+ 5.8
G	+ 1.5
H	+10.5
I	−14.2
J	+ 0.4
K	−49.7
L	− 6.0
M	−15.4
N	−10.9
	Mean = − 6.357

and

$$t = -1.487$$

From Table V, for 13 degrees of freedom and $t = -1.487$, we have $P = 0.17$, approximately. The value $P = 0.17$ is well above the critical level, 0.05. Both tests indicate that one type of pipe is not more liable to corrosion than the other.

Fieldner and Selvig (13) give the following data on the ash content of dry coal. Each pair of samples came from a different coal supply.

Sample A	Sample B	Sample A	Sample B	Sample A	Sample B
8.91	9.02	13.04	13.08	12.82	12.79
11.47	11.36	12.75	12.23	9.87	9.69
9.81	10.63	11.52	11.65	8.85	9.22
9.34	9.44	10.03	10.21	10.49	10.58
9.73	9.88	10.75	10.06	9.16	9.39
10.22	10.03	9.77	10.16	11.35	11.72
8.35	10.26	11.90	12.11	12.29	12.43
10.19	10.20	13.66	13.08	7.95	7.44
11.49	11.45	12.94	13.12	9.14	9.77
13.20	12.95	12.36	12.83	9.32	10.01
13.73	14.42	5.89	5.75	4.16	4.08
11.51	11.21	6.22	5.99	8.41	8.72
10.60	10.60	5.27	5.36	5.70	6.01
11.11	10.94	5.69	5.91	4.43	4.40
10.39	10.05	5.47	5.33	4.69	4.52
10.59	11.20	5.05	4.93	4.51	4.50
9.88	9.87	5.25	5.37	3.42	3.32
11.18	11.51	12.66	13.01	3.87	3.77
10.58	11.27	12.12	12.56	4.25	4.06

It is clear that the pairs of values are positively correlated, coal source being the control. The samples each weigh 3 pounds and are prepared in identical fashion, so any differences between sample A and sample B are expected to be negligibly slight. Do the data support this expectation?

We form differences:

Sample A — Sample B	Sample A — Sample B	Sample A — Sample B
−0.11	−0.04	+0.03
+0.11	+0.52	+0.18
−0.82	−0.13	−0.37
−0.10	−0.18	−0.09
−0.15	+0.69	−0.23
+0.19	−0.39	−0.37
−1.91	−0.21	−0.14
−0.01	+0.58	+0.51
+0.04	−0.18	−0.63
+0.25	−0.47	−0.69
−0.69	+0.14	+0.08
+0.30	+0.23	−0.31
0.00	−0.09	−0.31
+0.17	−0.22	+0.03
+0.34	+0.14	+0.17
−0.61	+0.12	+0.01
+0.01	−0.12	+0.10
−0.33	−0.35	+0.10
−0.69	−0.44	+0.19

We then find

$$t = -1.972$$

From Table V, for $t = -1.972$ and 56 degrees of freedom P is slightly below 0.05. The sample means may be considered significantly different, although the margin is slight. We conclude that experimental technique is subject to improvement, or that the samples A and B differ with respect to a non-randomized factor.

If the exceptional deviate $d = -1.91$ (the seventh value) is omitted, the result is

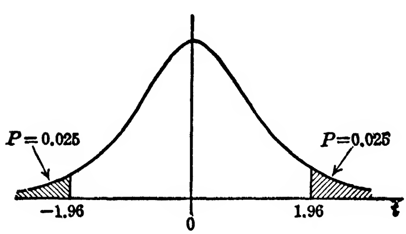
$$t = -1.679$$

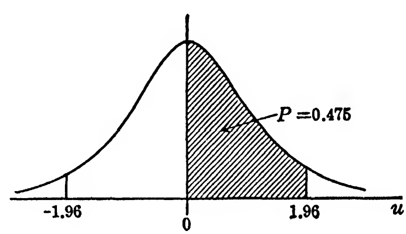
which, for 55 degrees of freedom gives $P > 0.05$ and the difference in mean ash content is judged not to be significant. Omission of an observation or observations is an unsound practice and should be done only when the investigator has reason to believe that the observation

in question was subject to special influences not affecting the remaining observations.

1.21 Large samples, paired variates. The estimate from small samples of the population variance is subject to error, and the amount of the error depends on the size of the samples. Hence in determining the probability P that a given value of $t = \bar{d}/\hat{\sigma}_{\bar{d}}$ could be exceeded in random sampling from a normal population we must take into account the number of independent observations (degrees of freedom) on which the estimate of the population variance is based. Accordingly, the probabilities of exceeding t given in Table V depend on the number of degrees of freedom.

If the experiment is relatively large, with, say, more than 30 observations in each sample, the population variance can be assumed to be given exactly by the sample variance, and the distribution of t passes into the normal distribution of u (areas under which are given in Table IV). The probability P that a given value of $u = \bar{d}/\sigma_{\bar{d}}$ could be exceeded does not involve the concept of degrees of freedom; this is evidenced by the absence of degrees of freedom in Table IV. The normal approximation to t is completely valid only if the sample size n is infinitely large; only the probabilities associated with the infinite sample sizes shown in the last row of Table V will correspond to those of Table IV. For example, in Table V for $n = \infty$ and $t = 1.96$, we find $P = 0.05$, as indicated in the illustration at the left.





From Table IV for $u = 1.96$ we find a value of 0.475; the area of the two tails is 0.05 as before; this is shown in the illustration at the right. If $n > 30$, the normal values given in Table IV may be safely used.

Let us reanalyze the data on the ash content of coal, now considering the sample of d_i to be large. We have

$$\bar{d}' = 0$$
$$\bar{d} = 0.1079$$
$$n = 57$$

$$\sigma^2 = \frac{9.553}{57} = 0.168 \quad \text{and} \quad \sigma = 0.41$$

$$\sigma_{\bar{d}} = \frac{0.41}{\sqrt{57}} = 0.054$$

$$u = \frac{\bar{d}}{\sigma_{\bar{d}}} = \frac{0.1079}{0.054} = -1.99$$

From Table IV, $P = 0.0466$. The difference is judged barely significant, as before.

1.22 Large samples, unpaired variates. If the observations in two small samples are, or are considered to be, unpaired, the estimate $\hat{\sigma}^2$ of the variance of the normal population is

$$\frac{\sum(X - \bar{X})^2 + \sum(Y - \bar{Y})^2}{n_X + n_Y - 2} \equiv \left(\frac{n_X s_X^2 + n_Y s_Y^2}{n_X + n_Y - 2}\right)$$

and the distribution of the statistic

$$t = \frac{\bar{d}}{\hat{\sigma}_{\bar{d}}} = \frac{\bar{d}}{\hat{\sigma}\sqrt{\dfrac{1}{n_X} + \dfrac{1}{n_Y}}}$$

is known but is not normal. If, however, the two samples are large, say $n_X > 30$, $n_Y > 30$, the population variance may be assumed to be known and to be given by the weighted mean of sample variances, i.e.,

[4] $$\sigma^2 = \frac{n_X s_X^2 + n_Y s_Y^2}{n_X + n_Y}$$

and the statistic t, which we have called u under these conditions, may be written

$$u = \frac{\bar{d}}{\sigma_{\bar{d}}} = \frac{\bar{d}}{\sigma\sqrt{\dfrac{1}{n_X} + \dfrac{1}{n_Y}}}$$

where σ is the square root of [4]; u is distributed normally.

1.23 Examples of 1.22. The British Cotton Industry Research Association (5) records the following results of breaking load tests on two types of yarn:

Type of yarn	Mean breaking load in ounces	Standard deviation	Number in sample
X	6.83	1.23	1782
Y	7.48	1.33	1914

Do the yarns differ significantly in their mean values? We have

$$\sigma_{\bar{d}}^2 = \frac{n_X s_X^2 + n_Y s_Y^2}{n_X + n_Y}\left(\frac{1}{n_X} + \frac{1}{n_Y}\right) = \frac{s_X^2}{n_Y} + \frac{s_Y^2}{n_X}$$

$$= \frac{(1.23)^2}{1914} + \frac{(1.33)^2}{1782}$$

$$= 0.001783$$

$$\frac{\bar{d}}{\sigma_{\bar{d}}} = \frac{0.65}{0.04223} = 15.39$$

This deviation is so improbable that it cannot be located in Table IV. Hence the yarns differ significantly in their means. The difference is, however, only 0.65 ounce and may be of slight practical significance.

Van Rest (43) gives the following data and calculations on the effect of stain (outdoor storage) on the hardness and bending strength of wood.

	Hardness		Bending Strength	
	Stained	*Unstained*	*Stained*	*Unstained*
Number of tests	40	100	40	100
Mean	117	132	6,184	6,270
Sum of squares about mean	8,655	27,244	16,799,390	30,459,499

Are hardness and bending strength significantly affected by stain? Previous formulae yield

$$\sigma_{\bar{d}}^2 = \frac{\sum(X - \bar{X})^2 + \sum(Y - \bar{Y})^2}{n_X \cdot n_Y}$$

we obtain:

$$\text{Hardness } \sigma_{\bar{d}} = 2.996 \quad \text{and} \quad \frac{\bar{d}}{\sigma_{\bar{d}}} = 5.007$$

$$\text{Bending strength } \sigma_{\bar{d}} = 108.7 \quad \text{and} \quad \frac{\bar{d}}{\sigma_{\bar{d}}} = 0.791$$

Hardness is really affected by stain, whereas bending strength is not, for the probabilities from Table IV are respectively $P = 0.0000003$ (highly significant) and $P = 0.2148$ (not significant).

1.24 Examples in which the hypothetical mean is not zero. Frequently in industrial practice we may want to use as the population mean the mean of a large number of observations of an earlier date or a figure set by a standards-making body. The practical problem is to

determine whether or not the mean of the current sample differs significantly from such a population mean.

The first illustration deals with small samples. Beckwith (2) gives the following data on the pile wool content, in ounces per three-quarters of a yard, of a fabric.

26.0
27.2
26.5
26.8
27.0

Quality specifications require a mean of 27.4. Are the data of this small sample compatible with the hypothesis that the mean of the population from which the sample was drawn is 27.4?

The appropriate method of analysis — which has already been used — is summarized as follows: If a quality characteristic X is normally distributed with mean $\bar{X}'$ and unknown variance σ^2, the quantity $\dfrac{\bar{X} - \bar{X}'}{\hat{\sigma}_{\bar{X}}}$ is distributed as Student's t with $n - 1$ degrees of freedom, where $\hat{\sigma}_{\bar{X}} = \sqrt{\hat{\sigma}^2_{\bar{X}}}$, and

$$\hat{\sigma}^2_{\bar{X}} = \frac{\sigma^2}{n}$$

and the best estimate $\hat{\sigma}^2$ of the unknown variance σ^2 from a single sample is

$$\frac{\sum (X - \bar{X})^2}{n - 1}$$

A test of normality of the population would have to be based on five observations; it will not be attempted. We have

$$t = \frac{\bar{X} - \bar{X}'}{\sigma_{\bar{X}}} = \frac{\bar{X} - \bar{X}'}{\hat{\sigma}/\sqrt{n}} = -3.34$$

For $n - 1$, i.e., for four degrees of freedom and for $t = -3.34$ we find $P = 0.015$ (15 samples in 1000). As this is a very low probability, the material at hand must be considered significantly different from the specification in its mean.

The following example illustrates the case for large samples: Pettebone and Young (32) record the following 1306 readings on the heat value in

Btu of a mixed gas. The data cover a period from January 1932 to January 1937.

Btu	Midpoints	Number of days
548.5–550.5	549.5	6
546.5–548.5	547.5	3
544.5–546.5	545.5	6
542.5–544.5	543.5	30
540.5–542.5	541.5	57
538.5–540.5	539.5	118
536.5–538.5	537.5	202
534.5–536.5	535.5	260
532.5–534.5	533.5	284
530.5–532.5	531.5	197
528.5–530.5	529.5	103
526.5–528.5	527.5	36
524.5–526.5	525.5	3
522.5–524.5	523.5	0
520.5–522.5	521.5	1
		1306

On 64 days at irregular intervals in the 5-year period, state inspection was conducted. The 64 observations which constituted an apparently random sample from the population of 1306 observations are given in the following table.

Btu (*midpoints*)	Number of days
549.5	1
547.5	1
545.5	3
543.5	3
541.5	5
539.5	10
537.5	11
535.5	9
533.5	8
531.5	6
529.5	5
527.5	0
525.5	1
523.5	0
521.5	1
	64

So far as means are concerned, is it likely that this constitutes a random sample from the given population?

The appropriate procedure is summarized as follows: If a quality characteristic X is distributed normally with mean $\bar{X}'$ and variance σ^2,

the means of random samples each of n observations will be distributed normally with mean $\bar{X} = \bar{X}'$ and variance σ^2/n.

In the present example a test of the normality of the population is hardly necessary, for it will presently be noted that if a sample is larger than 50 and if the population is as much as 10 times as large as the sample, the tendency to normality of the distribution of the means of random samples is negligibly affected by the nature of the population. If a test is to be applied, the statistics a and $\sqrt{b_1}$ or b_2 and $\sqrt{b_1}$ are computed. When testing the normality of the parent population from a small sample, the statistic a is better than b_2. In the present example 1306 observations are available, and we shall use the more familiar b_2 test. For a normal population

$$\beta_2 = \frac{\dfrac{\sum^{N}(X - \bar{X}')^4}{N}}{\left[\dfrac{\sum^{N}(X - \bar{X}'^2)}{N}\right]^2} = 3$$

The distribution of

$$b_2 = \frac{\dfrac{\sum^{n}(X - \bar{X})^4}{n}}{\left[\dfrac{\sum^{n}(X - \bar{X})^2}{n}\right]^2}.$$

is known and the 5 per cent and 1 per cent values are given in Table II. For our data

$$\sqrt{b_1} = 0.43$$

$$b_2 = 3.58$$

which indicate that the population is not normal, for both tests yield probabilities of less than 0.01. Normality of the *means* can, however, be assumed. We have

$$\bar{X}' = 534.99$$
$$\sigma = 3.85$$

$$\bar{X} = 536.72$$
$$n = 64$$

Also

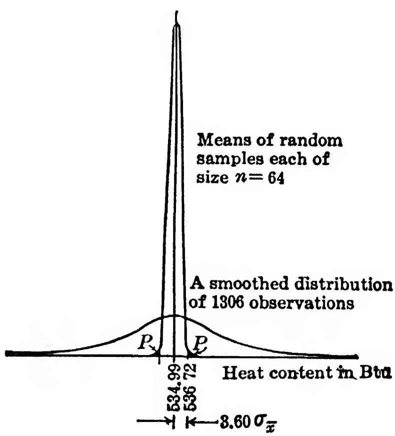
$$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} = \frac{(3.85)^2}{64} = 0.2316$$
$$\sigma_{\bar{X}} = 0.481$$

To enter Table IV, form

$$u = \frac{\bar{X} - \bar{X}'}{\sigma_{\bar{X}}} = \frac{1.73}{0.483} = +3.60$$

From Table IV we find that only 4 times in 10,000 trials would 3.60 be exceeded if chance alone is responsible for the deviation. This is a very small probability; therefore the mean of the inspector's readings departs significantly from the population mean, and reasons for this fact should be sought.

The problem and its solution is shown in the following illustration. Each shaded area P is 0.0002.



NOTES

1.25 The $\sqrt{b_1}$ and b_2 tests for normality, in detail. The following example illustrates in detail the $\sqrt{b_1}$ and b_2 tests for normality. In connection with the example we shall show certain short-cut methods of calculating the mean and the variance.

Pulsifer (33) gives the following data on the tensile strength, in actual load pounds, of 1000 cap screws of a certain dimension.

Tensile strength in pounds	Number of screws	Tensile strength in pounds	Number of screws
15,500	1	17,800	41
600	..	900	29
700	6	18,000	42
800	8	100	49
900	4	200	48
16,000	11	300	41
100	6	400	33
200	15	500	48
300	11	600	52
400	18	700	28
500	5	800	48
600	10	900	27
700	19	19,000	35
800	23	100	25
900	19	200	15
17,000	20	300	15
100	23	400	8
200	36	500	3
300	33	600	3
400	35	700	1
500	31	800	2
600	33	900	..
700	39	20,000	1
			1000

Is the population of tensile strengths normally distributed?

The normal population distribution is given by

$$y = \frac{f}{N} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\bar{X}'}{\sigma}\right)^2} \tag{5}$$

where y is the fractional frequency $\left(\frac{\text{frequency } f}{\text{total frequency } N}\right)$ or probability of screws with tensile strength X, $\bar{X}'$ is the mean tensile strength, and σ is the standard deviation. For [5] the values of $\sqrt{\beta_1}$ and β_2 for $N \rightarrow \infty$ are respectively 0 and 3, where

$$\sqrt{\beta_1} = \frac{\nu_3}{\nu_2^{3/2}}$$

and

$$\beta_2 = \frac{\nu_4}{\nu_2^2}$$

ν_k being defined as the kth moment of the deviation $x = X - \bar{X}'$, i.e.,

$$\nu_k = \int_{-\infty}^{\infty} x^k \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} dx$$

Estimates of $\bar{X}'$, σ, $\sqrt{\beta_1}$ and β_2 from the data of large samples are

$$\begin{aligned}
\bar{X} \quad (\rightarrow \bar{X}') &= \frac{\sum^{n} X}{n} \\
s^2 \quad (\rightarrow \sigma^2) &= \frac{\sum^{n} (X - \bar{X})^2}{n} \quad (= \mu_2) \\
\sqrt{b_1} \quad (\rightarrow \sqrt{\beta_1}) &= \mu_3/\mu_2^{3/2} \\
b_2 \quad (\rightarrow \beta_2) &= \mu_4/\mu_2^2
\end{aligned}$$

where

$$\mu_k \ (\rightarrow \nu_k) = \frac{\sum^{n} (X - \bar{X})^k}{n}$$

and n is the number of observations in the large sample.

We regroup the data to facilitate computation, although a minor grouping error is thereby introduced and certain information is lost. A correction will later be introduced which will partially remove this error.

Tensile strength Class Midpoints	Number of Screws
15,500	1
800	18
16,100	32
400	34
700	52
17,000	62
300	104
600	103
900	112
18,200	138
500	133
800	103
19,100	75
400	26
700	6
20,000	1
	1000

To compute the moments μ_k it is convenient to substitute for X a new variable Z:

$$X = a + cZ$$

where a is an arbitrary constant and c is the interval between class midpoints; in our case $c = 300$. Summing this expression over all observations and dividing by n we obtain

$$\bar{X} = a + cD$$

where $D = \sum fZ/n$, f being the frequency of X (or Z). By substituting these relations into the equation for μ_k we obtain by simple algebraic expansion the following equations, which are more convenient for purposes of calculation:

$$n\mu_2 = c^2(\sum fZ^2 - nD^2)$$
$$n\mu_3 = c^3(\sum fZ^3 - 3D\sum fZ^2 + 2nD^3)$$
$$n\mu_4 = c^4(\sum fZ^4 - 4D\sum fZ^3 + 6D^2\sum fZ^2 - 3nD^4)$$

Notice that the class interval c will disappear in calculating $\sqrt{b_1}$ and b_2.

The computations are carried out in the following table. The last column, suggested by Charlier, is for check purposes, for

$$\sum f(Z+1)^4 = \sum fZ^4 + 4\sum fZ^3 + 6\sum fZ^2 + 4\sum fZ + n$$

X Tensile Strength Class Midpoints	f Number of Screws	Z Deviations in Class Interval Units from $a = 17{,}900$	fZ	fZ^2	fZ^3	fZ^4	$f(Z+1)^4$
15,500	1	−8	−8	64	−512	4,096	2,401
800	18	−7	−126	882	−6,174	43,218	23,328
16,100	32	−6	−192	1,152	−6,912	41,472	20,000
400	34	−5	−170	850	−4,250	21,250	8,704
700	52	−4	−208	832	−3,328	13,312	4,212
17,000	62	−3	−186	558	−1,674	5,022	992
300	104	−2	−208	416	−832	1,664	104
600	103	−1	−103	103	−103	103	......
900	112	0	......	....	......	......	112
18,200	138	+1	138	138	138	138	2,208
500	133	+2	266	532	1,064	2,128	10,773
800	103	+3	309	927	2,781	8,343	26,368
19,100	75	+4	300	1,200	4,800	19,200	46,875
400	26	+5	130	650	3,250	16,250	33,696
700	6	+6	36	216	1,296	7,776	14,406
20,000	1	+7	7	49	343	2,401	4,096
	1000		−1,201	8,569	−23,785	186,373	198,275
			+1,186		+13,672		
			−15		−10,113		

The Charlier check indicates that the calculations have probably been made correctly, for

$$\begin{aligned}198{,}275 &= 186{,}373 + 4(-10{,}113) + 6(8569) + 4(-15) + 1000\\ &= 198{,}275\end{aligned}$$

The moments are

$$\mu_2 = \frac{8569 - 1000\left(\dfrac{15}{1000}\right)^2}{1000}(300)^2 = 8.568775\,(300)^2$$

$$\mu_3 = \frac{-10{,}113 - 3\left(\dfrac{-15}{1000}\right)(8569) + 2(1000)\left(\dfrac{-15}{1000}\right)^3}{1000}(300)^3$$

$$= -9.727402\,(300)^3$$

$$\mu_4 =$$

$$\frac{186{,}373 - 4\left(\dfrac{-15}{1000}\right)(-10{,}113) + 6\left(\dfrac{-15}{1000}\right)^2(8569) - 3(1000)\left(\dfrac{-15}{1000}\right)^4}{1000}(300)^4$$

$$= 185.777788\,(300)^4$$

from which

$$\begin{aligned}\sqrt{b_1} &= 0.39\\ b_2 &= 2.53\end{aligned}$$

Are these values sufficiently close to 0 and 3? From Table I we find that fewer than one in one hundred samples of size $n = 1000$ would have $\sqrt{b_1}$ further than 0.39 from 0. For $b_2 = 2.53$, this probability is again <0.01, as shown in Table II. Hence the present sample cannot be assumed to have been randomly drawn from a normal population.

1.26 Correction of the moments. In estimating $\sqrt{\beta_1}$ and β_2 the moments μ_k may be corrected for errors resulting from grouping the original observations into classes. The error arises from the fact that we wish to estimate the values of $\sqrt{\beta_1}$ and β_2 of a continuous curve whereas our data form a discontinuous curve. The adjustments most generally applicable are those due to Sheppard (48). These adjustments assume that the corrected distribution has high order contact (very gradual tapering) with the X axis at its extremities.

$$\begin{aligned}\mu_2\text{ (corrected)} &= \mu_2 - \tfrac{1}{12}c^2\\ \mu_3\text{ (corrected)} &= \mu_3\\ \mu_4\text{ (corrected)} &= \mu_4 - \tfrac{1}{2}\mu_2 c^2 + \tfrac{7}{240}c^4\end{aligned}$$

For our data

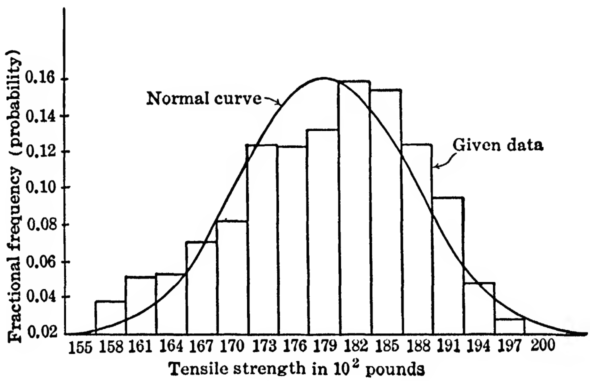
$$\begin{aligned}\mu_2\text{ (corrected)} &= 8.485442\,(300)^2\\ \mu_3\text{ (corrected)} &= -9.727402\,(300)^3\\ \mu_4\text{ (corrected)} &= 181.522567\,(300)^4\end{aligned}$$

and

$$\sqrt{b_1} = 0.39, \quad b_2 = 2.52$$

which yield the same conclusions.

The given data and the closest approximating normal distribution are shown in the following figure (actual fitting of a normal distribution to industrial data as a supplement to the $\sqrt{b_1}$ and b_2 test used here is seldom fruitful and the technique is not discussed here).



1.27 Some properties of the normal distribution. The X-variate of the normal distribution extends from $-\infty$ to $+\infty$; tensile strength, however, could not fall below zero. No serious error is introduced by this discrepancy.

In addition to properties already given, certain others may be noted. Writing $x = X - \bar{X}'$ we find, for a population of size N

$$\int_{-\infty}^{\infty} \frac{N}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}\, dx = N$$

that is, the area is N, the total frequency of observations. If we use the fractional frequency, i.e., probability y instead of $f = yN$, we find

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}\, dx = 1$$

The probability of x falling between $\pm\infty$ is unity; the probability of x falling between x_1 and x_2 is given by

$$\int_{x_1}^{x_2} \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}\, dx$$

Areas under this normal distribution (with abscissa x/σ) are given in Table IV. In that table x_1 is taken at zero.

The points of inflection for the normal curve occur at $x = \pm\sigma$. We have

$$\frac{dy}{dx} = -\frac{x}{\sigma^2}\frac{1}{\sigma\sqrt{2\pi}}e^{-x^2/2\sigma^2}$$

$$\frac{d^2y}{dx^2} = -\frac{1}{\sigma^3\sqrt{2\pi}}e^{-x^2/2\sigma^2} + \frac{x^2}{\sigma^5\sqrt{2\pi}}e^{-x^2 2/\sigma^2} = 0$$

from which

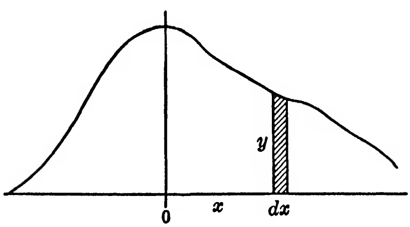
$$x = \pm\sigma$$

Other properties of the normal distribution are discussed later.

1.28 Moments. While the moments ν_k about the mean are used in this chapter in connection with normality, their definition is general. For a discontinuous distribution of n observations — with which we are always faced in practice — the kth moment about the mean has been defined by

$$\mu_k = \frac{\sum(X - \bar{X})^k}{n}$$

For a continuous population the corresponding definition of the kth moment ν_k about the mean is



$$\nu_k = \int_{-\infty}^{\infty} yx^k\,dx$$

where

$$x = X - \bar{X}'$$

1.29 That $\sqrt{\beta_1} = 0$ and $\beta_2 = 3$ for a normal distribution. A simple proof has been given by Bowley (3). The odd moments for the normal distribution are all zero; hence $\sqrt{\beta_1} = 0$. To prove, let ν_{2t+1} be any odd moment. Then

$$\nu_{2t+1} = \int_{-\infty}^{\infty} x^{2t+1}\frac{1}{\sigma\sqrt{2\pi}}e^{-x^2/2\sigma^2}\,dx$$

$$= \int_{-\infty}^{\infty}\varphi(x)dx$$

$$= \int_{0}^{\infty}\varphi(x)dx + \int_{-\infty}^{0}\varphi(x)dx$$

In the last term of the above substitute $-x' = x$. The limits become ∞ and 0.

We obtain

$$\nu_{2t+1} = \int_0^\infty \varphi(x)dx - \int_\infty^0 \varphi(-x')dx'$$

$$= \int_0^\infty \varphi(x)dx + \int_0^\infty \varphi(-x')dx' = 0$$

for $\varphi(x) = -\varphi(-x')$ in the function

$$\varphi(x)dx = x^{2t+1}\frac{1}{\sigma\sqrt{2\pi}}e^{-x^2/2\sigma^2}dx$$

This result would be obtained for any symmetrical distribution. Finally

$$\sqrt{\beta_1} = \frac{\nu_3}{\nu_2^{3/2}} = 0$$

for non-zero ν_2.

To determine the value of β_2 for a normal distribution, first consider ν_4.

$$\nu_4 = \int_{-\infty}^\infty x^4\frac{1}{\sigma\sqrt{2\pi}}e^{-x^2/2\sigma^2}\,dx$$

The solution is of the form

[6] $$x^3e^{-x^2/2\sigma^2}$$

for, after inclusion of the appropriate constants, the derivative of [6] yields the fourth and second moments.

Omitting constants, we have for the derivative

$$-\frac{1}{\sigma^2}x^4\,e^{-x^2/2\sigma^2} + 3x^2e^{-x^2/2\sigma^2}$$

Including the constants, we find

$$\left[\frac{\sigma}{\sqrt{2\pi}}x^3e^{-x^2/2\sigma^2}\right]_{-\infty}^\infty = -\int_{-\infty}^\infty \frac{1}{\sigma\sqrt{2\pi}}x^4e^{-x^2/2\sigma^2}\,dx$$

$$+\int_{-\infty}^\infty 3\sigma^2x^2\frac{1}{\sigma\sqrt{2\pi}}e^{-x^2/2\sigma^2}\,dx$$

or $$0 = -\nu_4 + 3\nu_2^2$$

or $$\beta_2 = \frac{\nu_4}{\nu_2^2} = 3$$

1.30 Short method of computing moments. Consider μ_2. We have

$$\mu_2 = \frac{\sum^{n}(X - \bar{X})^2}{n}$$

Make the substitutions $X = a + cZ$ and $\bar{X} = a + cD$ where

$$D = \frac{\sum Z}{n} \left(\equiv \frac{\sum fZ}{n} \text{ when the data are grouped} \right)$$

We obtain

$$\mu_2 = \frac{\sum(X - \bar{X})^2}{n} = \frac{\sum(a + cZ - a - cD)^2}{n}$$

$$= \frac{c^2\sum(Z - D)^2}{n}$$

$$= \frac{c^2(\sum Z^2 - 2\sum ZD + \sum D^2)}{n}$$

$$= \frac{c^2(\sum Z^2 - nD^2)}{n} = \frac{c^2}{n^2}[n \sum Z^2 - (\sum Z)^2]$$

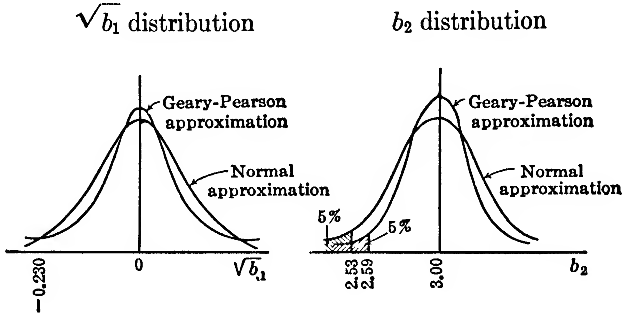
a form which facilitates rapid calculation. Similar forms have been given for μ_3 and μ_4.

1.31 Distributions of $\sqrt{b_1}$ and b_2. In order to determine whether or not the computed values of $\sqrt{b_1}$ and b_2 differ significantly from the normal population values $\sqrt{\beta_1} = 0$ and $\beta_2 = 3$, we need to be able to answer this question: If a large number of random samples each of size n are drawn from a population known to be normal and if the statistics $\sqrt{b_1}$ and b_2 are computed for each sample, what will be the distribution curve of $\sqrt{b_1}$ and that of b_2? The answer is not yet exactly known but the moments of the two distributions have been given by R. A. Fisher (15, *b*); on the basis of Fisher's results, E. Pearson and finally Geary and Pearson (20) have constructed approximate tables for various values of n; they have done similar work on a.

For very large n, the distributions of $\sqrt{b_1}$ and b_2 approach normality with standard deviations (usually called standard errors) respectively of $\sqrt{6/n}$ and $\sqrt{24/n}$, approximately. For $n < 1000$, the values given in Table I are to be preferred to normal approximations.

Assume that for $n = 300$ we have $\sqrt{b_1} = -0.230$. If this value is judged by reference to the areas given in Table I, we conclude that there are five chances in 100 that $\sqrt{b_1} = -0.230$ could have been exceeded (in a negative direction) in random sampling from a normal population. How does this compare with the normal approximation? The standard error of $\sqrt{b_1}$ is $\sqrt{6/300} = 0.1414$. Our deviation from the origin $\sqrt{b_1} = \sqrt{\beta_1} = 0$ is 0.230

or 1.626 standard error units; from the normal table the area to the left of $1.626\sigma_{\sqrt{b_1}}$ is 0.0520. Thus at this sample size and at this probability level the normal approximation to the distribution of $\sqrt{b_1}$ results in a slight overestimate of the proportion of samples with $\sqrt{b_1}$ to the left of -0.230.



Similar considerations hold for b_2, which is a test for flatness. For $n = 300$, $b_2 = 2.59$ (less peaked than the normal curve) is significantly different from 3.00 at the 5 per cent level (i.e., there is 1 chance in 20 that a sample of 300 observations drawn at random from a normal population ($\beta_2 = 3$) would have a value of b of 2.59 or less). Judged by the normal approximation $b_2 = 2.59$ is not quite significantly different from 3, for the area to the left of 2.59 will be found to be greater than 0.05.

In both instances the normal approximations lead to an overestimate of the proportion of large differences; in the case of b_2 the error is likely to be more serious, for while the Geary-Pearson approximation to the distribution of $\sqrt{b_1}$ is quite similar to a normal curve, the b_2 approximation differs appreciably.

In a normal population $\sqrt{\beta_1}$ and β_2 are independent of each other; hence they constitute independent tests of normality and for a large sample to be considered normal, both should be satisfied.

1.32 Outline of a derivation of the normal distribution. The importance of the normal distribution in sampling theory is evident. This distribution may originate in the following way: if a large number of independent causes, each producing a slight effect, affect a quality characteristic, values of the latter will, under certain conditions, be normally distributed. A derivation from Whittaker and Robinson (47) will be outlined.

The strength of cap screws varies from one screw to another. In other words, each shows a deviation from the average. This deviation will be assumed to be the effect of a large number of small deviations, the latter caused by the operation of a large number of independent causes, each of which has but a small effect.

Let the small deviations be

$$d_1, \cdots, d_n$$

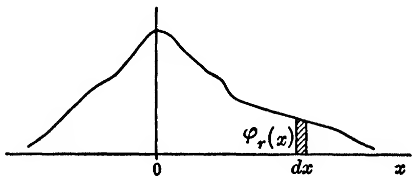
with the total effect or total deviation being

$$d_1 + \cdots + d_n$$

or more generally

[7] $$W_1 d_1 + \cdots + W_n d_n$$

where W_i are weights. What is the probability that for a given observation (strength of cap screw) this deviation will lie between α_1 and α_2? The probability that d_r lies between x and $x + dx$ is



$$\varphi_r(x)\ dx$$

The probability that d_r lies between d_1 and $d + dd_1 = \varphi_1(d_1)dd_1$; and the probability that d_r lies between d_2 and $d + dd_2 = \varphi_2(d_2)dd_2$, etc.

The probability of the concurrence of these deviations, if they are independent, is given by

[8] $$\varphi_1(d_1)\varphi_2(d_2) \cdots \varphi_n(d_n)dd_1dd_2 \cdots dd_n$$

Therefore the probability that [7] lies between α_1 and α_2 is the integral of [8] over all values satisfying

$$\alpha_1 < W_1 d_1 + \cdots + W_n d_n < \alpha_2$$

The integration leads to

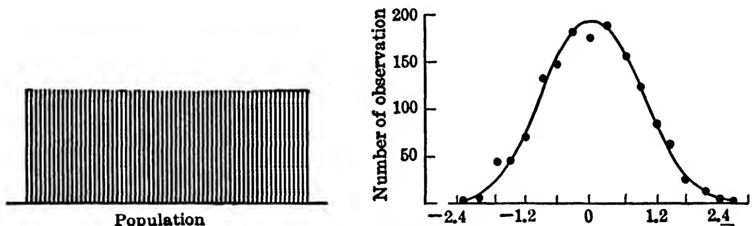
$$\varphi(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\theta x - (\theta^2/2!)I_2 + (i\theta^3/3!)I_3 + \cdots}\, d\theta$$

where the semi-invariants $I_2, I_3, \cdots, I_n$ are simple functions of the moments $\nu_2, \nu_3, \cdots, \nu_n$. If I_2 is finite and if most of the deviations $d_1 \cdots d_n$ are of the same order of magnitude, the higher semi-invariants $I_3 \cdots I_n$ will generally be small in comparison with I_2. Hence

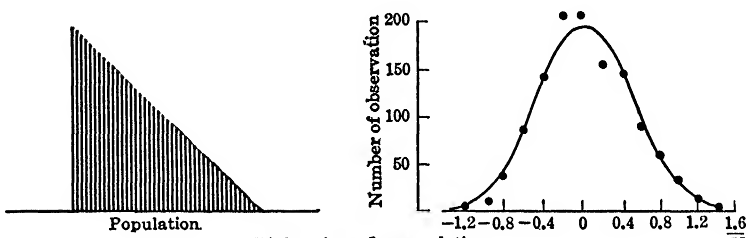
$$\begin{aligned}\varphi(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\theta x - (\theta^2/2)I_2}\, d\theta \\ &= \frac{1}{\sqrt{2\pi I_2}}\, e^{-x^2/2I_2} \\ &= \frac{1}{\sigma\sqrt{2\pi}}\, e^{-x^2/2\sigma^2}\end{aligned}$$

1.33 Normality of the distribution of means. Various investigations indicate that the distribution of means of random samples is approximately normal even when the samples are drawn from decidedly non-normal popu-

lations. For example, Shewhart (36) obtains the following striking results from drawing 1000 samples each of 4 items from rectangular and triangular populations. Normal curves have been fitted to the distributions of means.



Rectangular population



Right triangular population

Carver's students (11) have considered a population of the following non-normal character.

X	Frequency
3	2
15	9
29	43
405	189
1710	37

They found the distribution of 1000 means of random samples each of size 25 to be

X	Frequency
200—	2
280—	54
360—	203
440—	310
520—	254
600—	130
680—	36
760—	9
840—	2
	1000

Carver concluded from this and other results that if the sample size is 50 or more and the parent population is at least 10 times as large as the sample, the shape of the parent population has relatively slight control over the shape of the curve of means.

Exact and near-exact distributions of means have been found for various specific, non-normal populations and, in nearly all cases, the approach of the means to normality, even for low n, is evident. Thus, for a rectangular population Rietz and separately Irwin have found that the distribution of means rapidly approaches normality. This agrees with Shewhart's experimental results which have already been mentioned. The distribution of the means of samples drawn from a moderately skewed population known as Pearson's Type III has been found separately by Irwin, Church, and C. C. Craig. The result is another Type III distribution which rapidly approaches normality even for $n < 50$.

Using Craig's methods, Ness (29) found similar results for another non-normal population, Pearson's Type X. Baker and later Craig found distributions from still other non-normal populations; their results support the opinions stated above. The extensive literature on sampling from non-normal populations has been summarized by Rietz (35) and by Rider (34, *b*); their articles contain references to the mathematical work cited above with the exception of the unpublished results of Ness.

The proof of the normality of the curve of means when the samples are drawn from a normal population and the proof that the variance of the mean is given by σ^2/n will now be given.

1.34 Normality of the mean and the difference of two means. We first show that if x and y are independent and normally distributed about means of zero with variances respectively of σ_x^2 and σ_y^2, then $x + y$ (or $x - y$) is normally distributed with zero mean and with variance $\sigma_x^2 + \sigma_y^2$.

A procedure due to Jackson (23) will be used.

Given $\varphi(x,y)$, the frequency function for the joint distribution of x and y. To find $\varphi(u)$ where $u = x + y$.

The frequency function of a single variable may be found by integrating the joint frequency function over all possible values of the other variable. Thus

$$W(x) = \int_{-\infty}^{\infty} \varphi(x,y)dy$$

Consequently if $u = x + y$, then

$$\varphi(x,y) = \varphi(x,u - x)$$

and therefore

$$\psi(u) = \int_{-\infty}^{\infty} \varphi(x,u - x)dx$$

is the frequency function of the variable $u = x + y$.

If x and y are normally distributed about zero, i.e., if

$$\varphi_1(x) = c_1 e^{-ax^2}, \quad \varphi_2(y) = c_2 e^{-by^2}$$

$$a = \frac{1}{2\sigma_x^2}, \qquad b = \frac{1}{2\sigma_y^2}$$

we have, for x and y independent,

$$\varphi(x,y) = c_1 c_2 e^{-ax^2-by^2}$$

or

$$\psi(u) = c_1 c_2 \int_{-\infty}^{\infty} e^{-ax^2-b(u-x)^2}\, dx$$

Write

$$ax^2 + b(u-x)^2 = (a+b)\left(x - \frac{b}{a+b}u\right)^2 + \frac{ab}{a+b}u^2$$

and

$$x - \frac{b}{a+b}u = v, \qquad c = \frac{ab}{a+b}$$

We have

$$\psi(u) = c_1 c_2 e^{-cu^2} \int_{-\infty}^{\infty} e^{-(a+b)v^2}\, dv$$

The integration yields a constant multiplied by the entire area under the normal probability distribution (unity).

$$\psi(u) = c_0 e^{-cu^2}$$

$$= c_0 e^{-\frac{ab}{a+b}u^2} = c_0 e^{-\frac{u^2}{2(\sigma_x^2+\sigma_y^2)}}$$

which proves the theorem. The result for $x - y$ is the same.

The theorem may easily be generalized to n variables.

If $x_1, \cdots, x_n$ are independent and form a random sample from a normal population of variance σ^2, then $x_1 + \cdots + x_n$ is normally distributed with variance $n\sigma^2$. Also

$$\frac{x_1 + x_2 + \cdots + x_n}{n}$$

will be normally distributed with variance σ^2/n.

To prove the normality of the sum and of the mean of observations, write

$$x_1 + x_2 + x_3 = (x_1 + x_2) + x_3$$

Now the already proved theorem on the normality of the sum of two variables will apply to $(x_1 + x_2)$ and x_3. The extension to the sum or the mean of n variables is obvious.

1.35 Variance of the mean and the difference of means. As for the variances of the sum of n variables and of their mean it may be well to show that these relationships (and those already found on variances) are independent of theorems on normality. Let the variances of X and Y be respectively σ_x^2 and σ_y^2, where, for large samples

$$\sigma_x^2 = \frac{\sum f_x(X - \bar{X})^2}{n_x}, \qquad \sigma_y^2 = \frac{\sum f_y(Y - \bar{Y})^2}{n_y}$$

f_x and f_y being the frequencies of X and Y respectively. Frequently we write the variances in the form $\dfrac{\sum(X - \bar{X})^2}{n}$, the frequency f_x being implied though not explicitly introduced.

For continuous populations with means $\bar{X}$ and $\bar{Y}$ set equal to zero, these definitions are

$$\sigma_x^2 = \int_{-\infty}^{\infty} x^2\varphi(x)dx \qquad \sigma_y^2 = \int_{-\infty}^{\infty} y^2\psi(y)dy$$

where $x = X - \bar{X}$ and $y = Y - \bar{Y}$.

By definition of the variance we have for the variance of $x + y$ where x and y are independent

$$\sigma_{x+y}^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [(x + y) - 0]^2\varphi(x)\psi(y)dx\,dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2\varphi(x)\psi(y)dx\,dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y^2\varphi(x)\psi(y)dx\,dy$$

$$+ 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\varphi(x)\psi(y)dx\,dy$$

But

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\varphi(x)\psi(y)dx\,dy = \int_{-\infty}^{\infty} x\varphi(x)dx \int_{-\infty}^{\infty} y\psi(y)dy = 0$$

Therefore

$$\sigma_{x+y}^2 = \int_{-\infty}^{\infty} x^2\varphi(x)dx \int_{\infty}^{\infty} \psi(y)dy + \int_{-\infty}^{\infty} y^2\psi(y)dy \int_{-\infty}^{\infty} \varphi(x)dx$$

$$= \sigma_x^2 + \sigma_y^2$$

since

$$\int_{-\infty}^{\infty} \psi(y)dy = \int_{-\infty}^{\infty} \varphi(x)dx = 1$$

Note that

$$\sigma^2_{x-y} = \sigma^2_{x+y}$$

These results have already been reached in the special case of normal x and normal y. If x and y are from a population of variance σ^2,

$$\sigma^2_{x+y} = \sigma^2 + \sigma^2$$

or for n variables

$$\sigma^2_{\text{sum}} = n\sigma^2$$

For the variance of the mean of $x + y$

$$\sigma^2_{(x+y)/2} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{x+y}{2} - 0\right)^2 \varphi(x)\psi(y)dx\,dy$$

$$= \frac{\sigma_x^2 + \sigma_y^2}{4}$$

For the case in which x and y are from a population of variance σ^2

$$\sigma^2_{(x+y)/2} = \frac{\sigma^2}{2}$$

For the mean of n variables from a population of variance σ^2

$$\sigma^2_{\text{mean}} = \frac{\sigma^2_{x_1} + \cdots + \sigma^2_{x_n}}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

which was to be shown.

1.36 Mean estimate of σ^2. Given a population of kn observations divided so that there will be k samples with n observations in each sample, k being very large. We wish to form an estimate of the population variance σ^2, the unknown true value of σ^2 being

$$\frac{\sum^{k}\sum^{n}(X - \bar{X}')^2}{kn}$$

Write

$$\frac{\sum^{n}(X - \bar{X}')^2}{n} = \frac{\sum^{n}(X - \bar{X} + \bar{X} - \bar{X}')^2}{n}$$

$$= \frac{\sum^{n}(X - \bar{X})^2 + \sum^{n}(\bar{X} - \bar{X}')^2 + 2\sum^{n}(X - \bar{X})(\bar{X} - \bar{X}')}{n}$$

The cross product term is zero, for $\bar{X} - \bar{X}'$ is a constant and $\sum^{n}(X - \bar{X})$ is zero. Writing the variance of the sample as

$$\frac{\sum^{n}(X - \bar{X})^2}{n} = s^2$$

we have

$$\frac{\sum^{n}(X - \bar{X}')^2}{n} = s^2 + (\bar{X} - \bar{X}')^2$$

If this expression is summed over k samples and the result divided by k, we obtain

$$\frac{\sum^{k}\sum^{n}(X - \bar{X}')^2}{nk} = \frac{\sum^{k} s^2}{k} + \frac{\sum^{k}(\bar{X} - \bar{X}')^2}{k}$$

Write $\overline{s^2}$ for the mean sample variance $\sum s^2/k$. Now

$$\frac{\sum^{k}(\bar{X} - \bar{X}')^2}{k}$$

is seen to be the variance of the mean and is therefore equal to σ^2/n. We have

$$\sigma^2 = \overline{s^2} + \frac{\sigma^2}{n}$$

i.e., the "mean" estimate of σ^2 is

$$\frac{n}{n-1}\,\overline{s^2}$$

If σ^2 must be estimated from the data of a single sample, the estimate is

$$\frac{n}{n-1}\,s^2$$

which is equivalent to

$$\frac{\sum^{n}(X - \bar{X})^2}{n-1}$$

This estimate $\hat{\sigma}^2$ is thus shown to be the mean (unbiased) estimate of σ^2. It is also "best" in the sense that the variance of $\hat{\sigma}^2$ is a minimum, but this we do not show. It may be noted that while $\dfrac{\sum(X - \bar{X})^2}{n-1}$ is the mean estimate of σ^2, $\sqrt{\dfrac{\sum(X - \bar{X})^2}{n-1}}$ is not the mean estimate of σ; this fact is unimportant relative to the tests of significance used in this chapter.

1.37 Nature of the *t* test. It was shown that if we have a normal population of mean $\bar{X}'$ and variance σ^2 the means of random samples of size n are normally distributed with mean $\bar{X} = \bar{X}'$ and variance σ^2/n. Thus, if we reduce any deviation, say $\bar{X} - \bar{X}'$ to standard error units by forming

$$u = \frac{\bar{X} - \bar{X}'}{\sigma_{\bar{X}}} = \frac{\bar{X} - \bar{X}'}{\sigma/\sqrt{n}}$$

the probability of exceeding u is given by

$$\frac{1}{\sqrt{2\pi}}\int_u^\infty e^{-\frac{z^2}{2}}\,dz$$

Values of this probability integral may be found by subtracting the entries in Table IV from 0.5.

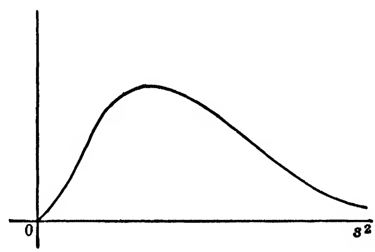
If σ^2 is unknown, the best estimate of σ^2 is

$$\hat{\sigma}^2 = \frac{n}{n-1}\,s^2$$

but

$$t = \frac{\bar{X} - \bar{X}'}{\hat{\sigma}/\sqrt{n}} = \frac{\bar{X} - \bar{X}'}{s/\sqrt{n-1}}$$

is not normally distributed, particularly not for small sample sizes, for when n is small, the standard deviation s varies considerably from sample to sample. The variability of s was discussed by earlier writers but to " Student " (39) belongs the credit both for recognizing the practical importance of the problem and for an approximately correct solution to the problem of the distribution of t.



To find the distribution of t, " Student " first found, by approximate methods, the distribution of s^2. He began by finding the first four moments of the distribution of s^2 in terms of the second moment σ^2 of the normal parent population.

The moments M_k of s^2 about the left end of the range ($s^2 = 0$) are found from simple expansions which yield, for examples

$$M_1 = \frac{\sum^k\sum^n(s^2 - 0)}{k} = \sigma^2\left(\frac{n-1}{n}\right)$$

$$M_2 = \frac{\sum^k\sum^n(s^2 - 0)^2}{k} = \sigma^4\,\frac{(n-1)(n+1)}{n^2}$$

Similar expressions may be found for the third and fourth moments M_3 and M_4 in terms of σ^2. These expressions are easily transformed to moments about the mean of s^2, and from these statistics the values of $\sqrt{b_1}$ and b_2 are computed. The values of $\sqrt{b_1}$ and b_2 indicate that a Pearson Type III curve will fit the distribution of s^2, from which the ordinate of the distribution of s^2 is found to be

$$y_{s^2} = c_1(s^2)^{(n-3)/2}\,e^{-ns^2/2\sigma^2}$$

and the ordinate of the distribution of s,

$$y_s = c_2 s^{n-2} e^{-ns^2/2\sigma^2}$$

c_1 and c_2 being constants.

"Student" then partially proved that $\bar{X}$ and s^2 were independent of each other. Thus, knowing the mean to be normally distributed and the standard deviation to be distributed as given above, he found the ordinate of the distribution of the ratio t to be

$$y_t = [\varphi(n)] \cdot \frac{1}{\left(1 + \dfrac{t^2}{n}\right)^{1/2\,(n+1)}}$$

a distribution which is symmetrical about $t = 0$, which is more peaked than the normal curve but which approaches normality as n becomes large. $\varphi(n)$ is known. Table V gives values of

$$2\int_t^{\infty} y_t dt$$

R. A. Fisher (15, *a*) later gave an exact proof of the distribution of t and of the complete independence of the mean and the variance of random samples drawn from a normal population.

1.38 Mean estimate of σ^2 from two small samples. If the quantity

$$\frac{\sum(X - \bar{X})^2 + \sum(Y - \bar{Y})^2}{n_X + n_Y - 2}$$

is summed over a large number, k, of samples and the sum divided by k, the resulting mean or "expected" value will be found to be equal to σ^2. This is easily demonstrated if we make use of three elementary properties of $E(X)$, the expected value of a variable X.

$$E(X) = \text{mean } X$$

$$E(cX) = c \text{ mean } X, \text{ where } c \text{ is a constant}$$

$$E(X + Y) = E(X) + E(Y)$$

We have

$$E\left[\frac{\sum(X - \bar{X})^2 + \sum(Y - \bar{Y})^2}{n_X + n_Y - 2}\right] = E\left[\frac{n_X s_X^2 + n_Y s_Y^2}{n_X + n_Y - 2}\right]$$

$$= \frac{1}{n_X + n_Y - 2}\,[n_X E(s_X^2) + n_Y E(s_Y^2)]$$

But

$$E(s_X^2) = \frac{n_X - 1}{n_X}\,\sigma^2 \quad \text{and} \quad E(s_Y^2) = \frac{n_Y - 1}{n_Y}\,\sigma^2$$

whence

$$E\left[\frac{\sum(X - \bar{X})^2 + \sum(Y - \bar{Y})^2}{n_X + n_Y - 2}\right] = \sigma^2$$

The mean value of the given function is σ^2. Hence the function is an unbiased estimate of σ^2. As in the case of estimating σ^2 from a single sample, the mean value given above is the "optimum" or best estimate of σ^2, in the sense that it has minimum variance.

1.39 " Student's " *t* applied to the difference of means. The application of the distribution of t to problems involving the difference of the means of two small samples arises from the fact that t is essentially the ratio of a normally distributed variable $\bar{X}$ to an independent estimate of the standard error of $\bar{X}$. Write

$$t^2 = \left(\frac{\bar{X} - \bar{X}'}{\hat{\sigma}/\sqrt{n}}\right)^2 = \left(\frac{\bar{X} - \bar{X}'}{s/\sqrt{n-1}}\right)^2$$

Replacing $\hat{\sigma}$ by $\sqrt{\dfrac{\sum(X - \bar{X})^2}{n-1}}$ and dividing numerator and denominator by σ^2 we obtain

$$t^2 = (n-1)\frac{\left(\dfrac{\bar{X} - \bar{X}'}{\sigma/\sqrt{n}}\right)^2}{\sum\left(\dfrac{X - \bar{X}}{\sigma}\right)^2}$$

The numerator $\dfrac{\bar{X} - \bar{X}'}{\sigma/\sqrt{n}}$ is normally distributed about zero with unit standard deviation; the denominator $\sum\left(\dfrac{X - \bar{X}}{\sigma}\right)^2$ is a function both of $\sum(X - \bar{X})^2$ and of the number of independent observations on which the estimate of σ^2 is based and its distribution is known. R. A. Fisher (15, *a*) was the first to note that any statistic which could be expressed as the ratio of a normally distributed variable to the square root of such an independently distributed estimate of the variance of that variable would be distributed as t with degrees of freedom equal to the number of independent observations from which the estimate of the variance was made.

This condition is satisfied in a difference of means test. If from normal populations of means $\bar{X}'$ and $\bar{Y}'$ and variance σ^2 we draw two random samples of sizes n_X and n_Y and means $\bar{X}$ and $\bar{Y}$ (optimum estimates of $\bar{X}'$ and $\bar{Y}'$), we already know that frequencies of $\bar{X}$ and $\bar{Y}$ are distributed normally about $\bar{X}'$ and $\bar{Y}'$ with respective variances of σ^2/n_X and σ^2/n_Y. We have shown that $\bar{X} - \bar{Y}$ is normally distributed about $\bar{X}' - \bar{Y}'$ with variance

$$\sigma^2\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)$$

The unbiased (and optimum) estimate of σ^2 is

$$\frac{\sum(X - \bar{X})^2 + \sum(Y - \bar{Y})^2}{n_X + n_Y - 2}$$

which is independent of $\bar{X} - \bar{Y}$. Hence

$$\frac{(\bar{X} - \bar{Y}) - (\bar{X}' - \bar{Y}')}{\sqrt{\dfrac{\sum(X - \bar{X})^2 + \sum(Y - \bar{Y})^2}{n_X + n_Y - 2} \cdot \dfrac{n_X + n_Y}{n_X n_Y}}}$$

is distributed as t with $n_1 + n_2 - 2$ degrees of freedom. In our examples and in general we are attempting to infer whether or not $\bar{X}' = \bar{Y}'$. Thus if $\bar{X}' - \bar{Y}' = 0$ lies beyond the 5 per cent level of t we conclude that the optimum estimates $\bar{X}$ and $\bar{Y}$ are significantly different, i.e., $\bar{X}' \neq \bar{Y}'$.

1.40 Correlation and the t test. Given

$$\begin{array}{cc} X_1 & Y_1 \\ X_2 & Y_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ X_n & Y_n \end{array}$$

We have tested the difference of means of small samples of X and Y in two ways. With unpaired variates we computed

$$\frac{\bar{X} - \bar{Y}}{\hat{\sigma}\sqrt{\dfrac{1}{n} + \dfrac{1}{n}}}$$

which is distributed as t with $2n - 2$ degrees of freedom. With paired variates, we formed

$$\begin{array}{l} X_1 - Y_1 = d_1 \\ X_2 - Y_2 = d_2 \\ \cdot \quad\ \cdot \quad\ \cdot \\ \cdot \quad\ \cdot \quad\ \cdot \\ \cdot \quad\ \cdot \quad\ \cdot \\ X_n - Y_n = d_n \end{array}$$

and then computed

$$\bar{X} - \bar{Y} = \bar{d}$$

and

$$\frac{\bar{d}}{\sigma_{\bar{d}}} = \frac{\bar{d}}{\hat{\sigma}_d/\sqrt{n}}$$

which is distributed as t with $n - 1$ degrees of freedom.

Certain features of the two situations are brought out in the following example. From two normal populations we draw the two following samples:

X	Y
6	7
4	5
8	9
3	4
9	10
$\bar{X} = 6$	$\bar{Y} = 7$

We find

$$t = \frac{\bar{X} - \bar{Y}}{\sigma\sqrt{\dfrac{1}{n_X} + \dfrac{1}{n_Y}}} = \frac{\bar{X} - \bar{Y}}{\sqrt{\dfrac{\sum(X - \bar{X})^2 + \sum(Y - \bar{Y})^2}{n_X + n_Y - 2}\left(\dfrac{1}{n_X} + \dfrac{1}{n_Y}\right)}}$$

$$= \frac{-1}{\sqrt{\dfrac{26 + 26}{8} \cdot \dfrac{2}{5}}} = -0.620$$

which for eight degrees of freedom yields $P = 0.55$; the difference is not significant.

If we apply the same method to the following data, X and Y having the same variates as in the preceding case,

X	Y
6	7
4	4
8	10
3	9
9	5
$\bar{X} = 6$	$\bar{Y} = 7$

we obtain exactly the same result. But the two sets of data differ strikingly. In the first set, whenever X is greater than $\bar{X}$, the Y paired with that X is greater than $\bar{Y}$ and whenever X is less than $\bar{X}$, the paired Y is less than $\bar{Y}$. In the second set of data, on the other hand, there appears to be little correlation between X and Y; for example, when X is greater than $\bar{X}$, Y is in one case greater than $\bar{Y}$ ($X = 8$, $Y = 10$) whereas in another case Y is less than $\bar{Y}$ ($X = 9$, $Y = 5$).

Now consider the value of the correlation coefficient r where

$$r = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{ns_Xs_Y}$$

for both of the above cases. The value of the numerator of r varies from $-\infty$ to $+\infty$ with the amount and nature (negative and positive) of the correlation

between X and Y. The effect of the denominator is to reduce this variation to the range -1 to $+1$ and to eliminate the effect on r of the units in which X and Y happen to be expressed. If the relationship between X and Y can be described perfectly by the linear regression function $Y_r = a + bX$ (see Ch. IV), then $r = +1$. Similarly, if the deviations $X - \bar{X}$ and $Y - \bar{Y}$ are independent of each other, i.e., if there is no correlation, we have $r = 0$.

These are only a few of the important properties of r, but for our present purpose, no other properties will be needed.

For the first set of data $r = +1$, whereas in the second set $r = 0$.

The first difference of means test clearly does not distinguish between cases in which X and Y are correlated and those in which no correlation is present.

If the test

$$\frac{\bar{d}}{\sigma_{\bar{d}}}$$

is applied, we obtain for the first set of data

$$\begin{array}{r} d \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \\ \hline \bar{d} = -1 \end{array}$$

or t is infinite, i.e., the mean difference $\bar{d} = -1$ is certainly significant. For the second set, in which $r = 0$, we find

$$t = \frac{\bar{d}}{\sqrt{\frac{\sum(d - \bar{d})^2}{(n-1)n}}} = \frac{-1}{\sqrt{\frac{52}{4 \cdot 5}}} = -0.620$$

exactly as before, but now with four instead of eight degrees of freedom.

In the case of positively correlated X and Y, elimination of the correlation by forming differences showed that the mean difference of -1 was highly significant; on the other hand, in the case of uncorrelated X and Y the same test was less sensitive than the ordinary difference of means test, for we obtained the same value of t with a loss of four degrees of freedom.

We may note that the ordinary difference of means test may be modified so that it is equivalent to the second test. In place of

$$\frac{\sum(X - \bar{X})^2 + \sum(Y - \bar{Y})^2}{2(n-2)}$$

as an estimate of σ^2 in the original test, we use

[9] $$\frac{\sum(X - \bar{X})^2 + \sum(Y - \bar{Y})^2 - 2\sum(X - \bar{X})(Y - \bar{Y})}{2(n-2)},$$

an estimate of σ^2 based on $n - 1$ degrees of freedom for the cross product reduces the number of independent observations by $n - 1$. For the first set of data, $\sum(X - \bar{X})(Y - \bar{Y}) = 26$, we would obtain

$$\hat{\sigma}^2 = \frac{26 + 26 - 2(26)}{8} \cdot \frac{2}{5} = 0$$

or t is infinite for four degrees of freedom, which corresponds to the result obtained when the correlation was eliminated by the alternative method of forming $X_i - Y_i$.

It is not possible to say, in advance of actual trial, which of the two tests

$$t = \frac{\bar{X} - \bar{Y}}{\hat{\sigma}\sqrt{\frac{1}{n} + \frac{1}{n}}}, \qquad 2n - 2 \quad \text{degrees of freedom}$$

$$t = \frac{\bar{d}}{\sigma_{\bar{d}}}, \qquad n - 1 \quad \text{degrees of freedom}$$

will be the more sensitive in a paired experiment. If the variables X and Y are positively correlated, either forming differences or using [9] as an estimate of σ^2 in the ordinary difference of means test will reduce the variance of (and hence increase the significance of) the difference $\bar{X} - \bar{Y}$. This gain may be nullified by the loss of half of the original number of independent observations. In our examples this loss (from eight to four degrees of freedom) is of no importance because σ^2 declined from 26 to 0; this is, of course, an extreme example. If the variates are unpaired in the original experiment, the second method is not available.

R. A. Fisher (15, *a*) has summed up the matter in the following sentence: "When both methods are available, sometimes the one and sometimes the other is the more sensitive; if either shows a significant deviation, its testimony cannot be ignored."

CHAPTER II

DIFFERENCES AMONG SEVERAL MEANS

2.1 Example of a simple experimental arrangement. An industrial experimenter wishes to compare the effects of five types of grids, A, B, C, D, and E, on the vacuum of radio tubes. With each type of grid he uses five tubes. The results, expressed in terms of a measure of vacuum, follow

A	B	C	D	E
93.6	95.3	94.5	96.8	94.6
95.3	96.9	97.0	98.2	97.8
97.0	95.8	97.8	97.2	98.0
93.7	97.3	97.0	97.2	95.0
98.0	97.7	98.3	97.9	98.9

2.2 General nature of the analysis. Even if the five types of grids are the same, the five column means, i.e., grid means, are not likely to be identical. For if from five populations (which we shall assume to be normal) which are alike in their means as well as in their variances, five random samples each of five observations are drawn, the five sample means will differ among themselves by chance. Our problem is to determine whether or not the observed variation in column means can be so explained. If it cannot, the hypothesis that the five normal populations are alike in their means and variances is rejected. Then, if it can be shown that the data do not refute those parts of the hypothesis covering normality and equality of variances, it will be concluded that the means of the five populations differ significantly among themselves, i.e., the five types of grids differ significantly, in a statistical sense, in their effects on vacuum.

If the five types of grids are alike in their effects on vacuum, the column means will vary about their mean by an amount which is determinable from the variation of the individual observations in the columns about their respective column means. For if the only unidentifiable factor (differences among grids) is without effect, both variations among column means and within columns are allocable to the same host of unidentifiable factors. Notice that it is not stated that, if grids are

alike in their effects, the variation among column means will be *equal* to that within columns for, though both variations are caused by the same forces, means will practically always vary less than the individual observations of which they are formed.

If differences among grids really affect vacuum, this calculable relationship between variation among column means and variation within columns does not exist. For while variation within columns is still caused only by unidentifiable causes, variation among column means is now attributable to these factors *and* to real differences among grids. This brings us to the nature of the test of significance, later to be called the F test. First, interpreting the unallocable variation within columns as the error of the experiment, we can set limits on the amount of variation that would be expected among the column means — if the same host of unidentifiable factors affect both. If the observed variation among means is outside these limits, the hypothesis that the grids are without effect is rejected.

2.3 Randomization of factors. In the experimental arrangement described in 2.1, all factors which might affect vacuum (other than grids) must be allocated at random. For example, assume that several sealing machines are used. If all tubes with grid A are sealed by the first machine and all tubes with grid B are sealed by the second machine, any conclusions regarding the effect on vacuum of differences among grids are vitiated, for the observed differences among column means are allocable to machines and/or to grids. Such vitiation can be precluded by assigning machines to grids at random. This applies to all influential factors.

The experimental arrangement shown in 2.1 will be called a completely randomized arrangement.

2.4 Magnitude of the error. The error of the completely randomized experiment can be taken to consist of variation in vacuum unexplained by differences among the grids. This variation is made up of the effects of differences among sealing machines, personnel, etc., and is directly measured by the variation of observations *within* columns, for such variation does not involve differences among grids. Thus, if several sealing machines are used on the five tubes containing grid type A, the variation of the observations in the first column about the mean of the first column is partly the result of differences among machines. If the operators used on the machines are of different skills, the result will be still greater variation among the observations on vacuum within each column, that is, still larger experimental error.

2.5 Complete control. Experimental error can always be reduced by holding constant all factors except the one under investigation. If only one sealing machine is used for all 25 tubes, differences among

sealing machines no longer contribute to the error of the experiment. And if only one operator is used, the same is true of differences among operators.

It is often impossible to achieve, simultaneously, complete control over all influential factors. For example, in an experiment dealing with variation in the quality of yarn, the use of a single loom would prolong the experiment over many weeks and the experimental error, decreased by the absence of loom differences, would be increased by the influence of factors which change with the passage of time, such as workroom humidity, operator efficiency, etc. If this source of variation is reduced by conducting the experiment in a single week, many looms will be required and loom differences reenter. In any case, the types of experimental arrangements we shall now describe obviate the need for complete control; in addition, they can often be made to yield valuable information which cannot be obtained from completely controlled experiments.

2.6 Latin Square. The Latin Square is an arrangement which permits at least two factors (other than the one being studied) to vary during the experiment, and yet it excludes the principal component of their variation from the error of the experiment. Assume that sealing machines and operators are two factors which might affect vacuum. If five grids are to be compared, the Latin Square arrangement requires five machines and five operators. The machines and operators are allocated to grids (A, B, C, D, and E) in such a way that the separate grids, machines, and operators are associated in the same trio only once.

		Machine				
		1	2	3	4	5
Operator	1	E	B	D	A	C
	2	C	D	B	E	A
	3	A	C	E	B	D
	4	D	E	A	C	B
	5	B	A	C	D	E

In order to appreciate the merits of this arrangement, consider the completely randomized experiment. In that experiment, two types of variation were noted, namely, variation among the grid means and the unallocable variation of the individual observations about their respec-

tive grid means, i.e., variation within grids. There is no other source of variation in that experiment; the total variation, i.e., the variation of the 25 observations about the grand mean is made up of these two variations.

Now assume that the earlier data were obtained from the following Latin Square arrangement.

		Machine				
		1	2	3	4	5
Operator	1	E 98.0	B 95.8	D 97.2	A 97.0	C 97.8
	2	C 98.3	D 97.9	B 97.7	E 98.9	A 98.0
	3	A 93.6	C 94.5	E 94.6	B 95.3	D 96.8
	4	D 97.2	E 95.0	A 93.7	C 97.0	B 97.3
	5	B 96.9	A 95.3	C 97.0	D 98.2	E 97.8

The total variation is the same as before. The grid means are unchanged; hence the variation among grids is the same as before. If we subtract variation among grids from the total variation, we obtain a term, say b, which must be numerically equal to the error term in the randomized arrangement, i.e., the term called variation within grids. But while the latter represented variation unallocable to any specific factor or factors, the b of the Latin Square can be divided into three parts, two of which are allocable to specific factors and the third of which is unallocable, i.e., unidentifiable.

In the present example the two new identifiable factors are machines and operators. The variation due to differences among machines (variation among column means), which contributed heavily to the error of the completely randomized experiment, is removable from the error of the present experiment; for inasmuch as each grid and each operator have been used an equal number of times (once) with each machine, the removal of the effect of machine differences cannot vitiate the comparison of grids (or of operators). In fact, in a Latin Square it is not possible to attribute the mean effect of any one factor to either or both of the remaining two factors. The effects of the three factors are com-

pletely separated; each effect is measurable and removable without interference with the others.

In the completely randomized experiment, it was not possible to remove machine effects for, first, there was no stated record as to which machine was used with each grid and operator, and second, even if there were such a record, it is unlikely unless deliberately planned that just five machines would be used and that each would be used exactly once with each grid and each operator. If these conditions are not satisfied, machine effects cannot be removed. For example, assume that the completely randomized experiment was conducted as follows (the numbers in parentheses refer to the different machines):

Grid				
A	B	C	D	E
93.6 (1)	95.3 (2)	94.5 (5)	96.8 (3)	94.6 (5)
95.3 (1)	96.9 (4)	97.0 (1)	98.2 (3)	97.8 (5)
97.0 (3)	95.8 (4)	97.8 (1)	97.2 (2)	98.0 (5)
93.7 (3)	97.3 (2)	97.0 (1)	97.2 (4)	95.0 (4)
98.0 (3)	97.7 (2)	98.3 (5)	97.9 (2)	98.9 (4)

Exactly five machines were used, but the machine effect cannot be removed for such a step would in part remove any effect of grids. For example, the difference in the means of the first and second machines is especially entwined with the differences of grids B and C.

Returning to the Latin Square, it is apparent that if the effects of machine and operator differences are statistically significant the experimental error of the square (error in the sense of unexplainable variation) will be less than that of the completely randomized arrangement. In the notation of the following table b_3 will be less than b.

Comparable Indexes of Variation

Completely randomized experiment	Latin Square
Variation among grids (a) Variation within grids (b)	Variation among grids (a) $\left.\begin{array}{l}\text{Variation among machines } b_1\\ \text{Variation among operators } b_2\\ \text{Unallocable variation } b_3\end{array}\right\}(b)$
Total variation (c)	Total variation (c)

2.7 Size of a Latin Square. It is disadvantageous to use many machines and many operators for the Latin Square excludes only the variation among the row and column *means* from experimental error.

If many machines are used, the error will be large even if machine means are identical. The same is true for operators. This limit on the number of machines and operators automatically limits the number of types of grids which can be compared in a single square. A 10 by 10 square may be taken as the maximum.

A small Latin Square is unreliable for while a Square of any size tends to reduce experimental error, the error of our estimate of the true value of that error from, say, the nine observations in a 3 by 3 Square is high. If as few as three or four grids are to be compared, more than one Latin Square must be used.

2.8 Other considerations. In the Latin Square arrangement, more than two influential factors can be admitted and their effects on error can be excluded. Assume that along with machine and operator differences it is believed that workroom humidity at the time of sealing affects vacuum. The various humidities that actually occur may be divided into five classes; the first machine is used *only* at the lowest humidity, the second *only* at another humidity, etc. What was in the original Latin Square variation allocable to machines is now variation allocable to machines or humidity or both. As we no longer know exactly what causes this variation, there has been some loss of information. However, if the exclusive purpose of the experiment is to distinguish among grids, this loss is of no importance and there may be a gain in the form of a further reduction of error by the exclusion of the effects of a third factor, humidity.

Superior arrangements for handling more than two factors, such as Graeco-Latin Squares, will not be discussed here.

The Latin Square is an experimental arrangement in which the allocation of machines and operators is subject to a double restriction: Each machine must be used once with each grid and each operator; also each operator must be used once with each grid and each machine. Many Squares satisfy these requirements (for example, the rows in a Latin Square may be interchanged). The Squares actually used can be selected at random by many card-drawing schemes, which the reader can easily arrange for himself.

2.9 Randomized blocks. The Latin Square arrangement excludes the effect of at least two of the factors, say machines and personnel, from the unexplained variation to which differences in the third factor, grids, are compared. Assume that it was known that one of the two factors was without effect, for example, that sealing machines do not differ among themselves in their effect on vacuum. Only the effect of differences among operators need be excluded, and the following plan, known to agronomists as a randomized block arrangement, is appropriate.

Operator									
1		2		3		4		5	
95.8	(B)	98.0	(A)	96.8	(D)	95.0	(E)	95.3	(A)
97.0	(A)	98.3	(C)	94.6	(E)	97.3	(B)	97.0	(C)
97.8	(C)	97.9	(D)	93.6	(A)	93 7	(A)	96.9	(B)
97.2	(D)	97.7	(B)	95.3	(B)	97.0	(C)	97.8	(E)
98.0	(E)	98.9	(E)	94.5	(C)	97.2	(D)	98.2	(D)

Differences among the means of the blocks, i.e., among the operator means, can be removed for each type of grid (the only factor remaining) as each type of grid is represented once in each block. Similarly, grid differences can be removed, for each operator is represented once with each type of grid. Other factors, such as machines and humidity, are handled as in 2.8 or are allocated to grids strictly at random in order to prevent possible vitiation.

2.10 Analysis of variance in a completely randomized experiment. We shall now consider the method of analysis to be applied to the completely randomized experiment. If vacuum is unaffected by grid differences, any variation among the five grid means is caused by the same unidentifiable factors that cause variation of the individual observation " within " each grid. Now conceive of the observations within the first column as having been drawn from one normal population, the observations within the second column from a second normal population, and so on, and the grid means as having been formed from samples whose items were drawn from a sixth normal population. If these normal populations are identical, the six estimates of their common variance tend to equality. All within-column variation is, however, of the same nature, so we shall reduce the number of estimates to one pooled within-column estimate and one estimate based on variation among means.

How are these estimates formed? If we pool the within-grid variations for two columns (grids), an unbiased estimate of the population variance σ^2 is

$$\frac{\sum(X_1 - \bar{X}_1)^2 + \sum(X_2 - \bar{X}_2)^2}{n_1 + n_2 - 2}$$

as was proved in 1.38. The proof given there is easily extended to k columns; the unbiassed estimate of σ^2 formed from the pooled within-column variation of all k columns is

$$[1] \quad \hat{\sigma}_1^2 = \frac{\sum(X_1 - \bar{X}_1)^2 + \sum(X_2 - \bar{X}_2)^2 + \cdots + \sum(X_k - \bar{X}_k)^2}{n_1 + n_2 + \cdots + n_k - k}$$

In the notation used here X_1 is summed over the n_1 values of X in the first column, X_2 is summed over the n_2 values of X_2 in the second column, etc. In our example, $n_1 = n_2 = \cdots = n_k = 5$ and $k = 5$.

Now consider the estimate of σ^2 formed from the variation among column means. The k means $\bar{X}_1, \bar{X}_2, \cdots, \bar{X}_k$ constitute a sample of k variates; an unbiased estimate of the variance of the normal population of such variates, i.e., *of means*, is given by

$$[2] \qquad \frac{\sum_{c=1}^{k}(\bar{X}_c - \bar{X})^2}{k - 1}$$

But these variates are means, not individual observations, and it would be incorrect to expect [2] to be equal to [1]. [2] is an estimate of the variance of a population of means whose variance in terms of the variance of the individual observations has already been shown to be σ^2/n_c where n_c is the number of observations from which each mean is formed (in the present example, $n_c = 5$). Hence

$$[3] \qquad \hat{\sigma}_2^2 = \frac{\sum n_c(\bar{X}_c - \bar{X})^2}{k - 1}$$

is an unbiased estimate of σ^2.

A third unbiased estimate of σ^2 is obtained from all n observations taken together. This estimate is

$$[4] \qquad \hat{\sigma}_3^2 = \frac{\sum(X - \bar{X})^2}{n - 1}$$

It will be noted that the numerator of each estimate contains as many terms as there are observations in the experiment, (n). This is immediately apparent in [1] and [4] while in [3] there are k terms $(\bar{X}_c - \bar{X})^2$ and each is weighted by the number of observations n_c in the corresponding column, i.e., a total of n terms. It will simplify our terminology if we take advantage of this fact and understand the summation sign to include n terms. Thus [1] will henceforth be written

$$\frac{\sum(X - \bar{X}_c)^2}{n - k}$$

and [3] will be replaced by

$$\frac{\sum(\bar{X}_c - \bar{X})^2}{k - 1}$$

The results to this point can be summarized in the following table:

Source of variation	Sum of squares	Degrees of freedom	Mean square
Among k columns	$\sum(\bar{X}_c - \bar{X})^2$	$k - 1$	$\hat{\sigma}_1^2 = \dfrac{\sum(\bar{X}_c - \bar{X})^2}{k - 1}$
Within k columns	$\sum(X - \bar{X}_c)^2$	$n - k$	$\hat{\sigma}_2^2 = \dfrac{\sum(X - \bar{X}_c)^2}{n - k}$
Total	$\sum(X - \bar{X})^2$	$n - 1$	$\hat{\sigma}_3^2 = \dfrac{\sum(X - \bar{X})^2}{n - 1}$

Let X_{ij} represent the ith variate in the jth column, and let $\bar{X}_j$ represent the mean of the jth column. The sums of squares are

$$
\begin{aligned}
\sum(\bar{X}_c - \bar{X})^2 = {} & (\bar{X}_1 - \bar{X})^2 + (\bar{X}_1 - \bar{X})^2 + \cdots \\
& + (\bar{X}_1 - \bar{X})^2 \ (n_1 \text{ terms}) \\
& + (\bar{X}_2 - \bar{X})^2 + (\bar{X}_2 - \bar{X})^2 + \cdots \\
& + (\bar{X}_2 - \bar{X})^2 \ (n_2 \text{ terms}) \\
& + \cdots \\
& + (\bar{X}_k - \bar{X})^2 + (\bar{X}_k - \bar{X})^2 + \cdots \\
& + (\bar{X}_k - \bar{X})^2 \ (n_k \text{ terms})
\end{aligned}
$$

$$
\begin{aligned}
\sum(X - \bar{X}_c)^2 = {} & (X_{11} - \bar{X}_1)^2 + (X_{21} - \bar{X}_1)^2 + \cdots \\
& + (X_{n_11} - \bar{X}_1)^2 \\
& + (X_{12} - \bar{X}_2)^2 + (X_{22} - \bar{X}_2)^2 + \cdots \\
& + (X_{n_22} - \bar{X}_2)^2 \\
& + \cdots \\
& + (X_{1k} - \bar{X}_k)^2 + (X_{2k} - \bar{X}_k)^2 + \cdots \\
& + (X_{n_kk} - \bar{X}_k)^2
\end{aligned}
$$

$$
\begin{aligned}
\sum(X - \bar{X})^2 = {} & (X_{11} - \bar{X})^2 + (X_{21} - \bar{X})^2 + \cdots \\
& + (X_{n_11} - \bar{X})^2 \\
& + (X_{21} - \bar{X})^2 + (X_{22} - \bar{X})^2 + \cdots \\
& + (X_{n_22} - \bar{X})^2 \\
& + \cdots \\
& + (X_{1k} - \bar{X})^2 + (X_{2k} - \bar{X})^2 + \cdots \\
& + (X_{n_kk} - \bar{X})^2
\end{aligned}
$$

The most convenient forms for computing the various sums of squares are the following

$$\sum(\bar{X}_c - \bar{X})^2 = \sum\bar{X}_c^2 - \frac{(\sum X)^2}{n}$$

$$\sum(X - \bar{X}_c)^2 = \sum X^2 - \sum\bar{X}_c^2$$

$$\sum(X - \bar{X})^2 = \sum X^2 - \frac{(\sum X)^2}{n}$$

where $$\sum\bar{X}_c^2 = \frac{(\sum X_1)^2}{n_1} + \frac{(\sum X_2)^2}{n_2} + \cdots + \frac{(\sum X_k)^2}{n_k}$$

In the notes at the end of this chapter it is shown that the total sum of squares is equal to the sum of squares associated with among-column variation plus that associated with within-column variation.

The degrees of freedom are also additive. The number of degrees of freedom associated with each estimate will be given by the number of variates in each summation less the number of constants (means) about which the deviations of the variates are taken. Thus, for total variability, we have n values of X in

$$\sum(X - \bar{X})^2$$

less one mean, $\bar{X}$ (the mean of all the observations), or $n - 1$ degrees of freedom. For variability among columns, there are k values of $\bar{X}_c$ in

$$\sum(X_c - \bar{X})^2$$

less one mean $\bar{X}$ or $k - 1$ degrees of freedom. For within-column variability we have n values of X in

$$\sum(X - \bar{X}_c)^2$$

less k values of $\bar{X}_c$ or $n - k$ degrees of freedom.

2.11 The F test. Now let us review the hypothesis to which this analysis is relevant. The hypothesis H states that the k column means arise from identical normal populations, i.e., normal populations of the same mean $\bar{X}'$ and the same variance σ^2. If the hypothesis is true, the estimates $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$, and $\hat{\sigma}_3^2$ should be the same, within the allowable range of sampling error. If, however, the ratio of say $\hat{\sigma}_1^2$ to $\hat{\sigma}_2^2$ is significantly different from unity, the hypothesis must be rejected, i.e., the columns do not come from normal populations of the same mean and the same variance. Now if the ratio of $\hat{\sigma}_1^2$ to $\hat{\sigma}_2^2$ differs significantly from unity while (1) the a and $\sqrt{b_1}$ tests support normality (or normality is assumed) and (2) the L_1 test supports the hypothesis that all columns come from populations of the same variance it follows that the part of

H which is untenable is that the columns come from populations alike in their *means*; i.e., the column or grid means differ significantly among themselves in their effect on vacuum.

The value of the ratio F of any two estimates of σ^2 will tend to unity as the number of independent variates on which each estimate is based is increased. The distribution of F for random samples drawn from normal populations is known, i.e., the probability of obtaining a value of F larger than any given value is known. We compute the ratio of an estimate associated with a suspected " cause " to the estimate which best defines the error of the experiment. If the probability is small, say less than 0.05, that this ratio could have occurred by chance in sampling from normal populations of identical means and variances, the hypothesis that such were the populations is rejected.

The numerical analysis of the completely randomized experiment follows.

Source of variation	Sum of squares	Degrees of freedom	Mean square
Among grids	10.25	4	2.56
Within grids (error)	44.18	20	2.21
Total	54.43	24	....

$$F = \frac{2.56}{2.21} = 1.16$$

is based on 4 and 20 degrees of freedom. From Table VIII, F would have to be as large as 2.87 in order to overthrow the hypothesis. We conclude that grid differences are without effect on vacuum.

2.12 A t test after an F test. Had the entire set of grids differed significantly among themselves, the following procedure could have been used to determine whether or not the apparent best and second best grids differ significantly between themselves.

The estimate of the variance of a single observation is 2.21. The standard deviation is $\sqrt{2.21} = 1.49$. Each grid mean is based on five observations; the standard error of a grid mean is accordingly $1.49/\sqrt{5} = 0.6664$. The difference of any two grid means has a standard error $\hat{\sigma}_{\bar{d}}$

$$\hat{\sigma}_{\bar{d}} = \hat{\sigma}\sqrt{\frac{1}{n} + \frac{1}{n}} = 1.49\sqrt{\frac{2}{5}} = 0.9422$$

The difference of two means, to be significant, should exceed, say, $2.086\hat{\sigma}_{\bar{d}} = 1.97$, the figure 2.086 being at the 5 per cent level of t for

20 degrees of freedom. The actual difference of the two best grids is only 0.54. In fact no two grid means differ in their means by as much as 1.97.

If many grids are studied, two grid means may well differ "significantly" even though the F test indicates over-all homogeneity. For even in a homogeneous set of means, the difference between say the largest and smallest means will likely appear to be "significant." A t test applied to two means after over-all homogeneity has either been refuted or not must be used with caution.

2.13 Analysis of variance in a Latin Square. In the Latin Square let X represent an observation, $\bar{X}_g$ a grid mean, $\bar{X}_m$ a machine mean, $\bar{X}_o$ an operator mean, and $\bar{X}$ the grand mean. Let there be n observations, G grids, M machines, and O operators ($n = G^2$). Four independent estimates of the variance of the population can be found, and they are listed in the following table.

Source of variation	Sum of squares	Degrees of freedom	Mean square
Among grids	$\sum(\bar{X}_g - \bar{X})^2$	$G - 1$	$\sum\bar{X}_g - \bar{X})^2/G-1$
Among machines	$\sum(\bar{X}_m - \bar{X})^2$	$M - 1$	$\sum(\bar{X}_m - \bar{X})^2/M-1$
Among operators	$\sum(\bar{X}_o - \bar{X})^2$	$O - 1$	$\sum(\bar{X}_o - \bar{X})^2/O-1$
Residual (error)	U (obtained by subtraction)	V (obtained by subtraction)	U/V
Total	$\sum(X - \bar{X})^2$	$G^2 - 1$	....

For the data shown in 2.6 we have the table below.

Source of variation	Sum of squares	Degrees of freedom	Mean square
Among grids	10.25	4	2.56
Among machines	12.42	4	3.11
Among operators	29.59	4	7.40
Residual (error)	2.27	12	0.19
Total	54.53	24	....

Machines and operators account for a very large part of the variation which, in the completely randomized experiment, constituted error. We have, for grids

$$F = \frac{2.56}{0.19} = 13.47$$

which for 4 and 12 degrees of freedom is highly significant. The Latin Square arrangement shows that differences among the grids have a real

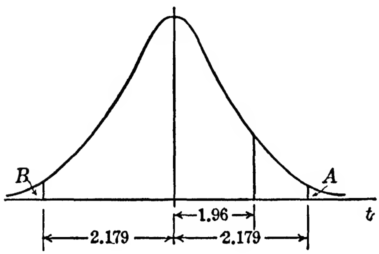
effect on vacuum, an effect which could not be determined from the completely randomized arrangement.

Is the difference between the means of the two best grids statistically significant? The variance of a single observation in the Latin Square arrangement is only 0.19, as against 2.21 in the completely randomized experiment. The standard deviation is $\sqrt{0.19} = 0.436$. We are interested in the standard error of the difference of two means, each based on five observations, and this is given by

$$0.436\sqrt{\tfrac{1}{5}+\tfrac{1}{5}}.$$

The observed difference in means of the two best grids is 0.54; or 1.96 standard error units, for

$$\frac{0.54}{0.436\sqrt{\frac{2}{5}}} = 1.96$$



This does not quite reach the 5 per cent value of t for 12 degrees of freedom, which is 2.179. Hence the observed difference between the two best grids cannot be said to be statistically significant. The accompanying graph illustrates this example; $A + B = 0.05$.

2.14 Analysis of variance in randomized blocks. The randomized block analysis is given in the table below:

Source of variation	Sum of squares	Degrees of freedom	Mean square
Among grids	10.25	4	2.56
Among operators	29.59	4	7.40
Residual (error)	14.69	16	0.92
Total	54.53	24	....

The effect of grids is not quite significant. This experiment has relative to grid differences a large experimental error (resulting from inclusion of machine effects in the error term), so that detection of differences in grid effects is not possible.

2.15 Further examples. Tippett (42, *b*) has described two textile experiments, one of which used randomized blocks and the other a Latin Square. The data from one of these experiments are shown immediately below but for the moment are assumed to come from a completely

randomized arrangement. The experiment in question was designed to determine the effect of differences in roller weightings on the strength of yarn. Three roller weightings A, B, and C were used and there were four strength tests for each weight. The quality of the yarn is measured by the product of lea strength in pounds and count.

A	B	C
1577	1535	1592
1690	1640	1652
1800	1783	1810
1642	1621	1663

The analysis of variance follows:

Source of variation	Sum of squares	Degrees of freedom	Mean square
Among weightings	3,000.8	2	1,500.4
Within weightings	83,982.2	9	9,331.4
Total	86,983.0	11	......

No test of significance is necessary, for F is less than unity. It must be concluded that the effect of differences in weighting is not statistically significant.

Actually this experiment employed a randomized block arrangement, the rows shown in the preceding table representing different sets of roving bobbins.

		Roller weighting		
		A	B	C
Roving set	1	1577	1535	1592
	2	1690	1640	1652
	3	1800	1783	1810
	4	1642	1621	1663

In the earlier description row differences were unallocable and variation among rows was an important element of the large experimental error. In the actual randomized block arrangement rows are identified as roving sets and, as each weighting is represented once in each roving

set, the roving effects can be removed without interfering with weighting effects. The result as seen below will be a sharply reduced experimental error.

Source of variation	Sum of squares	Degrees of freedom	Mean square
Among roller weights	3,000.8	2	1,500.4
Among roving sets	82,619.5	3	27,539.8
Residual (error)	1,362.7	6	227.1
Total	86,983.0	11	....

We find

$$F = \frac{1{,}500.4}{227.1} = 6.61$$

which for two and six degrees of freedom is statistically significant, $(P < 0.05)$. The evidence of this more sensitive experiment indicates that differences among roller weights do affect the strength of yarn.

The second experiment was designed to measure the effect of variations in sizing treatments on warp breakage rate. There are four treatments A, B, C, and D; there are two factors, loom and time differences, whose effects we wish to eliminate. For this problem a Latin Square arrangement is especially advantageous for, as already mentioned, neither of the two influential factors could be held constant throughout the experiment without augmenting the effect of the other on the error. If a single loom is used throughout, many time periods (weeks, approximately) will be needed, and variation associated with time will increase the error; whereas if the experiment is completed in a single week, many looms will be needed and loom differences will mount. The Latin Square arrangement effectively eliminates the principal effect of both sources of variation.

		Loom			
		1	2	3	4
Period	1	44 (D)	54 (A)	71 (C)	29 (B)
	2	22 (C)	59 (B)	100 (D)	22 (A)
	3	31 (A)	40 (C)	79 (B)	38 (D)
	4	27 (B)	83 (D)	100 (A)	29 (C)

The analysis of variance follows.

Source of variation	Sum of squares	Degrees of freedom	Mean square
Among looms	9,025.0	3	3,008
Among periods	370.5	3	124
Among sizes	1,389.5	3	463
Residual (error)	254.0	6	42
Total	11,039.0	15	....

We have

$$F = \frac{463}{42} = 11.02$$

which for three and six degrees of freedom is highly significant.

An F test applied to periods yields $P > 0.05$, i.e., the absence of a significant effect. Thus, while removal of the effect of loom differences clearly improved the precision of the experiment, the same is not true of time differences. Were such an experiment to be performed again, the question arises as to whether a randomized block arrangement should be used, for only one factor (looms) has a significant effect. With the present data, a randomized block arrangement (looms as in the Latin Square but periods allocated at random) would give the information shown in the accompanying table.

Source of variation	Sum of squares	Degrees of freedom	Mean square
Among looms	9,025.0	3	3008
Among sizes	1,389.5	3	463
Residual (error)	624.5	9	69.4
Total	11,039.0	15	......

$$F = \frac{463}{69.4} = 6.67$$

which yields $0.01 < P < 0.05$.

In this instance there is little to choose between the two arrangements, and inasmuch as we do not know in advance which factors are important, the Latin Square is preferable, at least for the original experiment. The advantage of the randomized block arrangement over the Latin Square is that the degrees of freedom wasted on an unimportant influence (periods) in the latter are allocated to error in the former. Thus the

error mean square of the block, while it happens to be larger, is more reliable, for it is based on nine rather than six independent variates, and the value of the ratio F needed to attribute significance to differences among sizings is smaller (3.86 as against 4.76). In the present example, this increase in the number of independent observations (from 6 to 9) is offset by the increase in the mean square (from 42 to 69.4). The effect of periods is not significant but it is much larger than the error of the Square, and this eliminates whatever advantage there might otherwise have been in a randomized block arrangement.

2.16 Other examples involving analysis of variance. Campbell and Lovell (6) give the following data on six independent sets of laboratory knock-ratings of a fuel.

Set 1	Set 2	Set 3	Set 4	Set 5	Set 6
70.5	69.7	70.5	71.4	71.0	69.5
71.9	70.5	70.7	70.5	71.3	70.6
71.0	70.4	71.0	71.2	70.8	71.5
71.5	70.2	70.5	70.8	70.7	70.6
71.1	71.0	70.3	70.1	69.8	70.2
70.1	71.0	71.2	70.8	70.5	70.6
69 8	71.4	70.1	71.4	70.6	71.0
70.5	70.5	71.0	71.0	70.0	70.8
70.0	70.8	70.4	71.0	69.9	71.4
71.1	70.9	70.0	70.6	70.8	70.1
....	70.5	71.1	70.6	....	70.2
....	71.2	71.0	70.4	....	....
....	....	71.4	....	....	....
....	....	71.0	....	....	....
....	....	71.0	....	....	....
....	....	71.2	....	....	....
....	....	70.4	....	....	....

Do the laboratory mean ratings differ significantly among themselves? Or may the six sets of ratings be combined?

The principal difference between this and preceding examples of completely randomized arrangements lies in the fact that the column means $\bar{X}_c$ are not equally important, for they are based on different numbers of observations. This fact affects somewhat the validity of the following analysis, but we shall assume this effect to be slight.

The within-sets sum of squares is

$$\sum(X - \bar{X}_c)^2$$
$$= \sum X^2 - \left[10\left(\frac{707.5}{10}\right)^2 + 12\left(\frac{848.1}{12}\right)^2 + \cdots + 11\left(\frac{776.5}{11}\right)^2\right]$$

and the among-sets term is

$$\sum(\bar{X}_c - \bar{X})^2 = \sum\bar{X}_c^2 - n\bar{X}^2$$
$$= \left[10\left(\frac{707.5}{10}\right)^2 + 12\left(\frac{848.1}{12}\right)^2 + \cdots + 11\left(\frac{776.5}{11}\right)^2\right]$$
$$- 72\left(\frac{5090.1}{72}\right)^2$$

The analysis of variance follows:

Source of variation	Sum of squares	Degrees of freedom	Mean square
Among sets	0.63	5	0.126
Within sets (error)	16.86	66	0.255
Total	17.49	71	.....

Actually, the mean square error is less (although not significantly less) than the mean square associated with variability among sets. No F test will be applied. The differences among the means of the six sets are not significant; the sets are homogeneous and they can be combined. The mean will be 70.7 and the standard error of the mean $\sqrt{(17.49/71)/72} = 0.059$.

If the data satisfy the assumptions underlying the method of analysis of variance, variation attributable to the action of specific factors and/or to their interaction can always be isolated. We shall give several examples.

Assume that two factors (and their interaction, see 2.17) are suspected of being responsible for variation. These three factors are isolated and the variation due to each is compared with the variation due only to experimental error; we thereby determine whether or not our suspicions are justified.

The data must provide a satisfactory estimate of experimental error and in the examples to be given this is true; for each value of the suspected causes (in the first example these causes are differences among lots and differences among rolls) there are several (three) values of the variable being studied (porosity). Within each set of three readings on porosity there is no change of lot or roll; whatever differences there are within each set of three readings are attributed to a large number of independent and unknown causes, each of which has a small effect. In short, these differences constitute experimental error.

Rider (34, a) gives the following Western Electric Company data on the porosity of condenser paper. Three readings are made on each of nine rolls from each lot.

		Reading number	Roll number								
			1	2	3	4	5	6	7	8	9
Lot number	I	1	1.5	1.5	2.7	3.0	3.4	2.1	2.0	3.0	5.1
		2	1.7	1.6	1.9	2.4	5.6	4.1	2.5	2.0	5.0
		3	1.6	1.7	2.0	2.6	5.6	4.6	2.8	1.9	4.0
	II	1	1.9	2.3	1.8	1.9	2.0	3.0	2.4	1.7	2.6
		2	1.5	2.4	2.9	3.5	1.9	2.6	2.0	1.5	4.3
		3	2.1	2.4	4.7	2.8	2.1	3.5	2.1	2.0	2.4
	III	1	2.5	3.2	1.4	7.8	3.2	1.9	2.0	1.1	2.1
		2	2.9	5.5	1.5	5.2	2.5	2.2	2.4	1.4	2.5
		3	3.3	7.1	3.4	5.0	4.0	3.1	3.7	4.1	1.9

The method of analysis of variance does not automatically suggest the appropriate breakdown of the data. Thus we might study the variance resulting from the differences between the means of rolls $\bar{X}_r$ (including all lots) and the grand mean $\bar{X}$, the corresponding sum of squares being

$$\sum(\bar{X}_r - \bar{X})^2$$

This would have slight value, for the position, say number 1, of a roll has no real meaning from lot to lot. The appropriate breakdown of the total sum of squares is

$$\underset{\text{Total}}{\sum(X - \bar{X})^2} = \underset{\substack{\text{Among}\\\text{lots}}}{\sum(\bar{X}_l - \bar{X})^2} + \underset{\substack{\text{Among rolls}\\\text{within lots}}}{\sum(\bar{X}_{rl} - \bar{X}_l)^2} + \underset{\substack{\text{Among measurements}\\\text{within lot-rolls}}}{\sum(X - \bar{X}_{rl})^2}$$

where $\bar{X}_l$ is the mean of a lot including all rolls, $\bar{X}_{rl}$ is the mean of a roll for a given lot, etc. Notice that, if the summation signs and squares are removed, the symbols "cancel out." Analysis of variance yields the following table:

Source of variation	Sum of squares	Degrees of freedom	Mean square
Among lots	7.90	2	3.95
Among rolls within lots	92.87	24	3.87
Among measurements within lot-rolls (error)	42.29	54	0.78
Total	143.06	80	....

As before, we may use as a practical rule the fact that each sum of squares has degrees of freedom equal to the number of variates summed less the number of independent relations between the variates. In the present example there are eighty-one observations so the variance estimate involving $\sum(X - \bar{X})^2$ will be based on eighty degrees of freedom, the mean $\bar{X}$ having been calculated from the observations. The among-

lots sum of squares $\sum(\bar{X}_l - \bar{X})^2$ involves three lot means less one relationship between them (again the grand mean); hence there are two degrees of freedom for the estimate based on this sum of squares. For among-rolls-within-lots, with sum of squares $\sum(\bar{X}_{rl} - \bar{X}_l)^2$ there are twenty-seven values of $\bar{X}_{rl}$ less three values of $\bar{X}_l$, or 24 degrees of freedom. In the among-measurements within-lot-rolls, sum of squares $\sum(X - \bar{X}_{rl})^2$, 81 values of X_i are summed but there are 27 relations among the X_i (27 roll-lot means $\bar{X}_{rl}$), leaving 54 degrees of freedom.

Another breakdown of the data would involve a split of the "among rolls within lots" sum of squares

$$\sum(\bar{X}_{rl} - \bar{X}_l)^2$$

into two terms, first, among rolls,

$$\sum(\bar{X}_r - \bar{X})^2$$

and a term which, as will be demonstrated, shows the joint or interaction effect of lots and rolls on porosity

$$\sum(\bar{X}_{rl} - \bar{X}_r - \bar{X}_l + \bar{X})^2$$

As in the previous breakdown, there will be 2 degrees of freedom for lots and 54 degrees of freedom for error. But the 24 degrees of freedom previously allocated to "among rolls within lots" must now be allocated to (*a*) among rolls and (*b*) interaction of lots and rolls. There are nine rolls and one restriction in the form of the grand mean, so the among-rolls estimate of the population variance is based on 8 degrees of freedom. By subtraction, 16 degrees of freedom are allocable to the interaction estimate of σ^2. Again notice the cancellation of symbols when the exponents and summation signs are removed.

The degrees of freedom allocated to the interaction (of lots and rolls) estimate will be shown at the end of this chapter to be the product of the degrees of freedom allocable to the constituent factors (among lots with 2 and among rolls with 8 degrees of freedom).

Analysis of variance yields the following table.

Source of variation	Sum of squares	Degrees of freedom	Mean square
Among lots	7.90	2	3.95
Among rolls	26.32	8	3.29
Interaction of rolls and lots	66.55	16	4.16
Error	42.29	54	0.78
Total	143.06	80	....

But differences among rolls for all lots have no practical significance whereas roll differences within lots are meaningful; hence the earlier breakdown of variability in quality is to be preferred.

The F test, as used in the analysis of variance, is essentially the ratio of variability associated with a suspected cause to error. For rolls (in the earlier breakdown) the appropriate ratio is

$$F_{\text{rolls}} = \frac{3.87}{0.78} = 4.96$$

From Table VIII, for 24 and 54 degrees of freedom, the 1 per cent level value of $F_{\text{rolls}} = 2.16$. The actual value of F exceeds this critical value. Variation in porosity is therefore partly attributable to differences among rolls within lots and, if possible, these differences should be eliminated.

In judging the variability among lots, the appropriate error sum of squares is 92.87 plus 42.29, with 24 plus 54 degrees of freedom, for both factors associated with these quantities clearly contribute to the error of comparing *lot means*. We have $F = \dfrac{3.95}{1.73} = 2.28$ which, from Table VIII, for 2 and 78 degrees of freedom, is not significant.

Rider (34, *a*) gives the following Western Electric Company data on impact strength, in foot-pounds, of specimens of insulating material. The specimens were cut lengthwise and crosswise from the sheets as indicated.

Type of cut		Specimen number	Lot number I	II	III	IV	V
Type of cut	Lengthwise specimens	1	1.15	1.16	0.79	0.96	0.49
		2	0.84	0.85	0.68	0.82	0.61
		3	0.88	1.00	0.64	0.98	0.59
		4	0.91	1.08	0.72	0.93	0.51
		5	0.86	0.80	0.63	0.81	0.53
		6	0.88	1.01	0.59	0.79	0.72
		7	0.92	1.14	0.81	0.79	0.67
		8	0.87	0.87	0.65	0.86	0.47
		9	0.93	0.97	0.64	0.84	0.44
		10	0.95	1.09	0.75	0.92	0.48
	Crosswise specimens	1	0.89	0.86	0.52	0.86	0.52
		2	0.69	1.17	0.52	1.06	0.53
		3	0.46	1.18	0.80	0.81	0.47
		4	0.85	1.32	0.64	0.97	0.47
		5	0.73	1.03	0.63	0.90	0.57
		6	0.67	0.84	0.58	0.93	0.54
		7	0.78	0.89	0.65	0.87	0.56
		8	0.77	0.84	0.60	0.88	0.55
		9	0.80	1.03	0.71	0.89	0.45
		10	0.79	1.06	0.59	0.82	0.60

The appropriate breakdown, written symbolically, follows:

Between types of cut $(\bar{X}_c - \bar{X})$

Among lots $(\bar{X}_l - \bar{X})$

Among specimens within a lot and within a type of cut, that is, experimental error

$$(X - \bar{X}_{lc})$$

The total variability is of the form $(X - \bar{X})$; we have

$$(X - \bar{X}) = (X_c - \bar{X}) + (X_l - \bar{X}) + (X - \bar{X}_{lc}) + R$$

from which R is of the form

$$(\bar{X}_{lc} - \bar{X}_c - \bar{X}_l + \bar{X})$$

a term which will presently be shown to symbolize the interaction or joint effect on impact strength of cuts and lots.

Source of variation	Sum of squares	Degrees of freedom	Mean square
Between types of cut	0.0454	1	0.0454
Among lots	2.7912	4	0.6978
Interaction of cuts and lots	0.1417	4	0.0354
Error	0.8947	90	0.0099
Total	3.8730	99	......

The cut-lot interaction sum of squares is best calculated by writing the following *totals*:

		Lot number					
		I	II	III	IV	V	Total
Cut	Lengthwise	9.19	9.97	6.90	8.70	5.51	40.27
	Crosswise	7.43	10.22	6.24	8.99	5.26	38.14
Total		16.62	20.19	13.14	17.69	10.77	78.41

We have for this table

$$\sum(X - \bar{X})^2 = \sum(\bar{X}_{\text{cut}} - \bar{X})^2 + \sum(\bar{X}_{\text{lot}} - \bar{X})^2 + \sum(X - \bar{X}_{\text{cut}} - \bar{X}_{\text{lot}} + \bar{X})^2$$

$$\sum(X - \bar{X})^2 = 10\left(\frac{9.19}{10}\right)^2 + 10\left(\frac{9.97}{10}\right)^2 + \cdots$$

$$+ 10\left(\frac{5.26}{10}\right)^2 - 100\left(\frac{78.41}{100}\right)^2 = 2.9783$$

$$\sum(\bar{X}_{\text{cut}} - \bar{X})^2 = 50\left(\frac{40.27}{50}\right)^2 + 50\left(\frac{38.14}{50}\right)^2 - 100\left(\frac{78.41}{100}\right)^2$$

$$= 0.0454$$

$$\sum(\bar{X}_{\text{lot}} - \bar{X})^2 = 2.7912 \quad \text{(already calculated)}$$

$$\text{Interaction sum of squares} = 2.9783 - 0.0454 - 2.7912 = 0.1417$$

Notice that the breakdown favored in this example is equivalent to the one not favored in the previous example. In the previous example, the column headings were roll numbers each of which was without meaning when taken over *all* lots. The number three roll, for instance, was the third roll chosen in each lot and if the selection is made at random, there is no reason to expect that the third roll should differ significantly from the others. In the present example, a column comprises one lot and lot differences are meaningful. Hence in the present example we are interested in

$$\sum(\bar{X}_{\text{column}} - \bar{X})^2$$

Finally

$$F_{\text{cut}} = \frac{0.0454}{0.0099} = 4.59; \qquad F_{\text{critical }(.05)} = 3.95$$

$$F_{\text{lot}} = \frac{0.6978}{0.0099} = 70.49; \qquad F_{\text{critical }(.05)} = 2.47$$

$$F_{\text{interaction}} = \frac{0.0354}{0.0099} = 3.58; \qquad F_{\text{critical }(.05)} = 2.47$$

Both cuts and lots, and their joint effect, are significantly responsible for variable quality.

As a final example, the following Western Electric Company data are given by Rider (34, *a*). They deal with the thickness of coating, in 0.0001 of an inch, on fibre strips sprayed with varnish. Measurements were taken at each of five different points on each of the three strips selected from each of five lots.

		Strip number	Point number 1	2	3	4	5
Lots	I	1	10	8	10	9	7
		2	8	8	8	8	10
		3	8	10	10	6	7
	II	1	13	12	12	12	13
		2	10	9	13	11	8
		3	11	8	10	12	12
	III	1	12	13	14	17	16
		2	17	10	13	10	14
		3	12	11	13	16	13
	IV	1	14	13	17	11	11
		2	11	9	13	11	12
		3	17	13	14	13	8
	V	1	9	13	17	13	11
		2	8	11	10	12	11
		3	7	14	14	9	9

When analyzing data which come from an experiment designed and carried out by someone other than the analyst, one must know what variation in the data is random variation. The term we call "error" would consist of variability in thickness among points for a given strip within a given lot if (*a*) the points were distributed at random and (*b*) the three strips were, say, consistently of three different kinds. If, however, the strips are randomly chosen from the lot while the different points refer systematically to certain parts of a strip, i.e., they are not randomly selected, the error term would better consist of the estimate of variance from the term among-strips. In the present example the former is true. The appropriate analysis is, therefore,

(*a*) variability among lots $(\bar{X}_l - \bar{X})$
(*b*) variability among strips within lots $(\bar{X}_{sl} - \bar{X}_l)$
(*c*) variability among points within strips $(X - \bar{X}_{sl})$

Symbolically

$$(X - \bar{X}) = (\bar{X}_l - \bar{X}) + (\bar{X}_{sl} - \bar{X}_l) + (X - \bar{X}_{sl}) + R$$

from which

$$R = 0$$

The $(X - \bar{X}_{sl})$ term is the experimental error term.

Source of variation	Sum of squares	Degrees of freedom	Mean square
Among lots	207.92	4	51.98
Among strips within lots	49.20	10	4.92
Among points within strips (error)	277.20	60	4.62
Total	534.32	74	.....

We have

$$F_{\text{lots}} = \frac{51.98}{4.66} = 11.15$$

and

$$F_{\text{strips}} = \frac{4.92}{4.62} = 1.06$$

the value 4.66 being the mean square compounded from among strips within lots and among points within strips (as in the example on page 72). The 5 per cent level values are 2.50 and 1.99. Only the lot variation is statistically significant. The association of strips with variation in coating thickness is not statistically significant.

2.17 Interaction. To illustrate the meaning of interaction, consider the two following examples (from Snedecor, 38).

		Column 1	Column 2	Column 3	Mean
Row	1	1.8	2.0	1.4	1.73
	2	1.6	1.8	1.2	1.53
	3	1.3	1.5	0.9	1.23
Mean		1.57	1.77	1.17	1.50

Analysis of variance yields the following table:

Source of variation	Sum of squares	Degrees of freedom	Mean square
Rows	0.38	2	0.19
Columns	0.56	2	0.28
Interaction of rows and columns	0	4	0
Total	0.94	8	....

where

$$\overset{\text{Total}}{\sum(\bar{X}_{rc}{}^{*}-\bar{X})^2}=\overset{\text{Rows}}{\sum(\bar{X}_r-\bar{X})^2}+\overset{\text{Columns}}{\sum(\bar{X}_c-\bar{X})^2}+\overset{\text{Interaction}}{\sum(\bar{X}_{rc}-\bar{X}_c-\bar{X}_r+\bar{X})^2}$$

The interaction sum of squares

$$\sum(X-\bar{X}_r-\bar{X}_c+\bar{X})^2$$

is given by

$$\begin{aligned}&(1.8-1.73-1.57+1.50)^2\\+&(1.6-1.53-1.57+1.50)^2\\+&\cdots\end{aligned}$$

each term of which is zero. To appreciate the meaning of zero interaction, notice that in proceeding from the first to the second column of the original data, all variates are increased by the same (absolute, not percentage) amount (0.2) and from the second to the third column all variates are decreased by the same amount (0.6), and similarly for rows. Variation among observations from column to column is the same regardless of which row is considered, i.e., there is no "interaction" between columns and rows.

As a second example, consider

		Column			Mean
		1	2	3	
Row	1	1.6	2.0	0.8	1.47
	2	1.5	1.0	1.9	1.47
	3	1.3	1.4	1.7	1.47
Mean		1.47	1.47	1.47	1.47

Analysis of variance yields the following table:

Source of variation	Sum of squares	Degrees of freedom	Mean square
Rows	0	2	0
Columns	0	2	0
Interaction of rows and columns	1.24	4	0.31
Total	1.24	8	....

* $\bar{X}_{rc} \equiv X$.

In this case, all variation is attributable to interaction. As one proceeds from column to column, the amount and direction of change of the variate is completely dependent on the row; thus from the first to the second column the variate increases by 0.4 in the first row, decreases by 0.5 in the second, and increases by 0.1 in the third. The algebraic sum of these changes is zero. Separately, the suspected "causes" (rows and columns) are responsible for none of the variability; operating jointly they are responsible for all of the variability.

Practical problems yield interactions somewhere between the extremes shown in these two examples. Finally, when more than two "causes" are under investigation (as in the following example) more complex interaction terms will be produced; these may be interpreted analogously to the above.

2.18 Formal analysis of variance. Data can be classified with respect to any number of factors (causes). In the few published examples dealing with four or more factors, certain effects are a priori considered unimportant and the analyst therefore uses a plan of analysis appropriate to his data but one which is rarely useful elsewhere.

			Pot 1			Pot 2		
		Journey	Cylinder			Cylinder		
			3	10	16	3	10	16
Run	I	1	47	56	100	52	61	88
		2	55	89	93	49	62	97
		3	35	57	56	34	60	72
		4	78	67	113	47	93	118
		5	33	40	128	16	29	130
	II	1	52	66	36	65	80	40
		2	21	61	49	122	97	79
		3	31	39	25	45	54	72
		4	43	72	52	109	120	80
		5	37	51	67	67	85	63
	III	1	50	61	60	75	139	130
		2	33	27	49	46	58	63
		3	24	39	24	15	33	39
		4	18	18	43	22	16	19
		5	28	42	28	27	19	22
	IV	1	24	34	43	46	66	24
		2	24	49	42	40	117	105
		3	21	21	51	30	28	34
		4	21	69	48	36	64	53
		5	76	48	42	39	60	78

Experience indicates that students have difficulty following such non-systematic procedures.

Examples involving multifold classification can always first be analyzed formally (systematically) and combining of terms can be left to the end. The following is an example of this procedure.

Tippett (42, *a*) gives the data in the previous table, from a paper by Gould and Hampton (21) on the mean number of seed (defects) per unit area of spectacle glass. Four factors (runs, journeys, cylinders, and pots) may affect the seed count; accordingly the experiment is conducted and the data are classified with respect to these factors. Cylinders of glass are made in pots, a journey is equivalent to a day, and glass made on consecutive days from the same pots constitutes a run. Three cylinders of the eighteen made (numbers 3, 10, and 16 in the order of manufacture) were studied.

These data can be broken down in many ways. Tippett uses the following breakdown, which yields mean squares having the greatest practical interest.

Source of variation	Degrees of freedom			
(a) Within pots				
(1) Between cylinders			16	
(2) Between journeys			32	
(3) Residual within pots			64	
(4) Total				112
(b) Between pots				
(5) Between runs			3	
(6) Residual between pots			4	
(7) Total				7
(c) Between cylinders				
(8) Common to all runs	2			
(9) Common to both pots in run less (8)	6			
(10) Specific to pot	8			
(11) Total		16		
(d) Between journeys				
(12) Common to all runs	4			
(13) Common to both pots in run less (12)	12			
(14) Specific to pot	16			
(15) Total		32		
Grand total				119

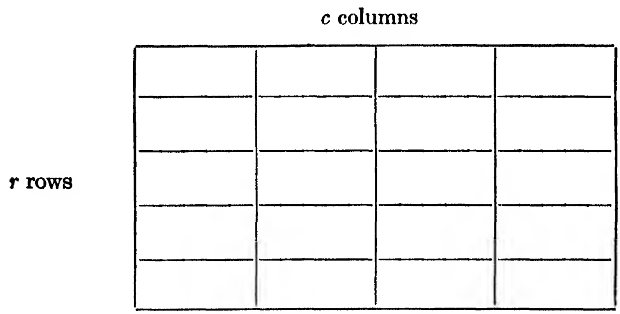
The original data provide no completely satisfactory index of experimental error for there is but one measurement for each combination of

suspected causes; it is the practice, particularly in unreplicated agricultural experimentation, to take a complex (uninterpretable) interaction term (interaction of pots, runs, journeys, and cylinders) as experimental error.

If we have two suspected causes (rows and columns), there are

$$C_1^2 + C_2^2 = 3$$

terms in a formal breakdown, C_1^2 representing the number of combinations of two things taken one at a time, etc.

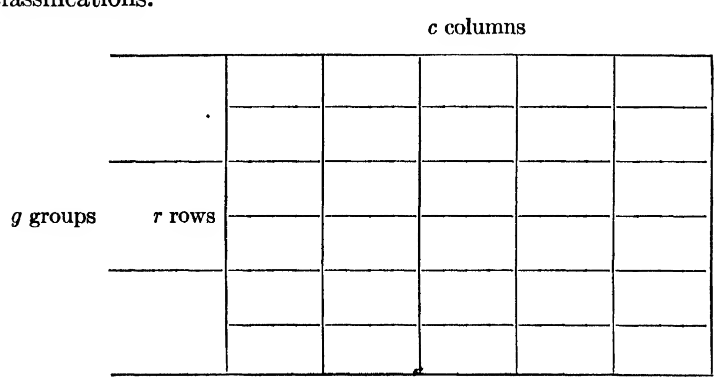


Source of variation	Sum of squares	Degrees of freedom
Among columns	$\sum(\bar{X}_c - \bar{X})^2$	$c - 1$
Among rows	$\sum(\bar{X}_r - \bar{X})^2$	$r - 1$
Interaction (columns $\times$ rows)	$\sum(X - \bar{X}_c - \bar{X}_r + \bar{X})^2$	$(c - 1)(r - 1)$
Total	$\sum(X - \bar{X})^2$	$cr - 1$

For a three-cause formal breakdown, there are

$$C_1^3 + C_2^3 + C_3^3 = 7$$

classifications.



Source of variation	Sum of squares	Degrees of freedom
Among columns	$\sum(\bar{X}_c - \bar{X})^2$	$c - 1$
Among rows	$\sum(\bar{X}_r - \bar{X})^2$	$r - 1$
Among groups	$\sum(\bar{X}_g - \bar{X})^2$	$g - 1$
First-order interaction (columns × rows)	$\sum(\bar{X}_{cr} - \bar{X}_c - \bar{X}_r + \bar{X})$	$(c - 1)(r - 1)$
First-order interaction (columns × groups)	$\sum(\bar{X}_{cg} - \bar{X}_c - \bar{X}_g + \bar{X})^2$	$(c - 1)(g - 1)$
First-order interaction (rows × groups)	$\sum(\bar{X}_{rg} - \bar{X}_r - \bar{X}_g + \bar{X})^2$	$(r - 1)(g - 1)$
Second-order interaction (columns × rows × groups)	$\sum(X - \bar{X}_{cr} - \bar{X}_{cg} - \bar{X}_{rg} + \bar{X}_c + \bar{X}_r + \bar{X}_g - \bar{X})^2$	$(c - 1)(r - 1)(g - 1)$
Total	$\sum(X - \bar{X})^2$	$crg - 1$

For a four-way formal breakdown, we have

$$C_1^4 + C_2^4 + C_3^4 + C_4^4 = 15$$

classifications. The numbers in parentheses show the degrees of freedom associated with the mean squares for the data of Gould and Hampton.

MAIN EFFECTS

Runs	$\sum(X_r - \bar{X})^2$	$(r - 1)$	(3)
Pots	$\sum(X_p - \bar{X})^2$	$(p - 1)$	(1)
Journeys	$\sum(X_j - \bar{X})^2$	$(j - 1)$	(4)
Cylinders	$\sum(X_c - \bar{X})^2$	$(c - 1)$	(2)

FIRST-ORDER INTERACTION

Runs × pots	$\sum(\bar{X}_{rp} - \bar{X}_r - \bar{X}_p + \bar{X})^2$	$(r - 1)(p - 1)$	(3)
Runs × journeys	$\sum(\bar{X}_{rj} - \bar{X}_r - \bar{X}_j + \bar{X})^2$	$(r - 1)(j - 1)$	(12)
Runs × cylinders	$\sum(\bar{X}_{rc} - \bar{X}_r - \bar{X}_c + \bar{X})^2$	$(r - 1)(c - 1)$	(6)
Pots × journeys	$\sum(\bar{X}_{pj} - \bar{X}_p - \bar{X}_j + \bar{X})^2$	$(p - 1)(j - 1)$	(4)
Pots × cylinders	$\sum(\bar{X}_{pc} - \bar{X}_p - \bar{X}_c + \bar{X})^2$	$(p - 1)(c - 1)$	(2)
Journeys × cylinders	$\sum(\bar{X}_{jc} - \bar{X}_j - \bar{X}_c + \bar{X})^2$	$(j - 1)(c - 1)$	(8)

SECOND-ORDER INTERACTION

Runs × journeys × cylinders

$\sum(\bar{X}_{rjc} - \bar{X}_{rj} - \bar{X}_{rc} - \bar{X}_{jc} + \bar{X}_r + \bar{X}_j + \bar{X}_c - \bar{X})^2$ $(r - 1)(j - 1)(c - 1)$ (24)

Runs × journeys × pots

$\sum(\bar{X}_{rjp} - \bar{X}_{rj} - \bar{X}_{rp} - \bar{X}_{jp} + \bar{X}_r + \bar{X}_j + \bar{X}_p - \bar{X})^2$ $(r - 1)(j - 1)(p - 1)$ (12)

Journeys × cylinders × pots

$\sum(\bar{X}_{jcp} - \bar{X}_{jc} - \bar{X}_{jp} - \bar{X}_{cp} + \bar{X}_j + \bar{X}_c + \bar{X}_p - \bar{X})^2$ $(j - 1)(c - 1)(p - 1)$ (8)

Runs × cylinders × pots

$\sum(\bar{X}_{rcp} - \bar{X}_{rc} - \bar{X}_{rp} - \bar{X}_{cp} + \bar{X}_r + \bar{X}_c + \bar{X}_p - \bar{X})^2$ $(r - 1)(c - 1)(p - 1)$ (6)

Third-order interaction

Runs × journeys × cylinders × pots

$$\sum(\bar{X}_{rjcp}{}^{*} - \bar{X}_{rjc} - \bar{X}_{rjp} - \bar{X}_{rjp} + \bar{X}_{rj} + \bar{X}_{rc} + \bar{X}_{jc} + \bar{X}_{rp} + \bar{X}_{jp} + \bar{X}_{cp} - \bar{X}_{r} - \bar{X}_{j} - \bar{X}_{c} - \bar{X}_{p} + \bar{X})^2 \qquad (r-1)(j-1)(c-1)(p-1) \quad (24)$$

Total $\sum(X - \bar{X})^2$ $\qquad rjcp - 1 \qquad (119)$

It is easier to compute and to appreciate the meanings of these terms if one sets down portions of the data in the three-cause form. Instead of writing the means in each case, the totals are used, for although we are always dealing with means in the form $\sum X/n$ in variance analysis, the means themselves need seldom be computed; $\sum X$ is sufficient and more convenient.

Table *a*

		Run			
		1	2	3	4
Journey	1	404	339	515	237
	2	445	429	276	377
	3	314	266	174	185
	4	516	476	136	291
	5	376	370	166	343

Table *b*

		Run			
		1	2	3	4
Cylinder	3	446	592	338	357
	10	614	725	452	556
	16	995	563	477	520

Table *c*

		Run			
		1	2	3	4
Pot	1	1047	702	544	613
	2	1008	1178	723	820

Table *d*

		Journey				
		1	2	3	4	5
Cylinder	3	411	390	235	374	323
	10	563	560	331	519	374
	16	521	577	373	526	558

Table *e*

		Journey				
		1	2	3	4	5
Pot	1	629	592	423	642	620
	2	866	935	516	777	635

Table *f*

		Cylinder		
		3	10	16
Pot	1	751	1006	1149
	2	982	1341	1406

* $\bar{X}_{rjcp} \equiv X$.

TABLE *g*

		Run 1			Run 2			Run 3			Run 4		
		Cylinder			Cylinder			Cylinder			Cylinder		
		3	10	16	3	10	16	3	10	16	3	10	16
Journey	1	99	117	188	117	146	76	125	200	190	70	100	67
	2	104	151	190	143	158	128	79	85	112	64	166	147
	3	69	117	128	76	93	97	39	72	63	51	49	85
	4	125	160	231	152	192	132	40	34	62	57	133	101
	5	49	69	258	104	136	130	55	61	50	115	108	120

TABLE *h*

		Run 1					Run 2				
		Journey					Journey				
		1	2	3	4	5	1	2	3	4	5
Pot	1	203	237	148	258	201	154	131	95	167	155
	2	201	208	166	258	175	185	298	171	309	215
		Run 3					Run 4				
		Journey					Journey				
		1	2	3	4	5	1	2	3	4	5
Pot	1	171	109	87	79	98	101	115	93	138	166
	2	344	167	87	57	68	136	262	92	153	177

TABLE *i*

		Pot 1			Pot 2		
		Cylinder			Cylinder		
		3	10	16	3	10	16
Journey	1	173	217	239	238	346	282
	2	133	226	233	257	334	344
	3	111	156	156	124	175	217
	4	160	226	256	214	293	270
	5	174	181	265	149	193	293

TABLE j

		Pot 1			Pot 2		
		Cylinder			Cylinder		
		3	10	16	3	10	16
Run	1	248	309	490	198	305	505
	2	184	289	229	408	436	334
	3	153	187	204	185	265	273
	4	166	221	226	191	335	294

Each table need not be completely analyzed, for there is a certain amount of duplication. Thus Table a will yield

Among runs $\sum(\bar{X}_r - \bar{X})^2$
Among journeys $\sum(\bar{X}_j - \bar{X})^2$
Interaction (runs × journeys) $\sum(\bar{X}_{rj} - \bar{X}_r - \bar{X}_j + \bar{X})^2$

and Table c will yield

Among runs $\sum(\bar{X}_r - \bar{X})^2$
Between pots $\sum(\bar{X}_p - \bar{X})^2$
Interaction (runs × pots) $\sum(\bar{X}_{rp} - \bar{X}_r - \bar{X}_p + \bar{X})^2$

duplicating "among-runs"; similarly for other tables. The formal four-way breakdown of the data is carried out in the table below.

Term number	Source of variation	Sum of squares	Degrees of freedom	Mean square
1	Runs	13,679.89	3	4,559.96
2	Journeys	9,684.00	4	2,421.00
3	Cylinders	9,132.87	2	4,566.44
4	Pots	5,644.41	1	5,644.41
5	Runs × journeys	18,650.07	12	1,554.17
6	Runs × cylinders	11,532.73	6	1,922.12
7	Runs × pots	4,455.16	3	1,485.05
8	Journeys × cylinders	1,992.55	8	249.07
9	Journeys × pots	2,727.13	4	681.78
10	Cylinders × pots	146.46	2	73.23
11	Runs × journeys × cylinders	9,104.18	24	379.34
12	Runs × journeys × pots	6,855.46	12	571.29
13	Journeys × cylinders × pots	917.12	8	114.64
14	Runs × cylinders × pots	1,320.47	6	220.08
15	Runs × journeys × cylinders × pots	6,384.29	24	266.01
	Total	102,226.79	119	

Practical interest would be expected primarily to center on the significance of the variability between pots, among runs, among journeys, and among cylinders. With this in mind how can we best classify the above fifteen terms? Note that an interaction term, such as, say term 9, can be classified beneath either of two headings—variability among journeys or between pots. "Between pots" has no interest, for the pots were not used in any particular sequence. Hence terms such as 9, involving pots and another factor, should be eventually listed under the other factor. In a similar way, "between runs" means little, for the runs are quite independent of each other. Finally, there is no question as to how the main effects (terms 1 to 4) are to be classified as only one factor is involved in each.

We obtain the following table:

Source of variation	Degrees of freedom
Among runs	
Runs (term 1)	3
Between pots	
Pots (term 4)	1
Among journeys	
Journeys (term 2)	4
Runs × journeys (term 5)	12
Journeys × pots (term 9)	4
Journeys × pots × runs (term 12)	12
Among cylinders	
Cylinders (term 3)	2
Runs × cylinders (term 6)	6
Cylinders × pots (term 10)	2
Runs × cylinders × pots (term 14)	6

The following terms are not automatically placed by the criterion of practical interest (run and pot variation of no interest):

	Degrees of freedom
Term 7 (runs × pots)	3
Term 8 (journeys × cylinder)	8
Term 11 (runs × journeys × cylinders)	24
Term 13 (journeys × cylinders × pots)	8
Term 15 (runs × journeys × cylinders × pots)	24

Of these the first will be arbitrarily classified as a between-pot term. The others are all journey × cylinder terms and constitute a residual

interaction, as will be seen from the following addition.

Term 8 $\bar{X}_{jc} - \bar{X}_j - \bar{X}_c + \bar{X}$

Term 11 $\bar{X}_{jrc} - \bar{X}_{jr} - \bar{X}_{jc} - \bar{X}_{rc} + \bar{X}_j + \bar{X}_r + \bar{X}_c - \bar{X}$

Term 13 $\bar{X}_{jcp} - \bar{X}_{jc} - \bar{X}_{jp} - \bar{X}_{cp} + \bar{X}_j + \bar{X}_c + \bar{X}_p - \bar{X}$

Term 15 $\bar{X}_{jcpr} - \bar{X}_{jcp} - \bar{X}_{jcr} - \bar{X}_{jpr} - \bar{X}_{cpr} + \bar{X}_{rc} + \bar{X}_{pr} + \bar{X}_{rj}$
$+ \bar{X}_{jc} + \bar{X}_{jp} + \bar{X}_{cp} - \bar{X}_j - \bar{X}_r - \bar{X}_c - \bar{X}_p + \bar{X}$

The encircled terms cancel out, leaving

$$\bar{X}_{jcpr} - \bar{X}_{jpr} - \bar{X}_{cpr} + \bar{X}_{pr}$$

which is the interaction term of $\bar{X}_{jpr}$ and $\bar{X}_{cpr}$, i.e., the joint effect on seed of journeys and cylinders, the effect differing from one pot-run to another. This is a secondary or residual effect and is classified as such.

Source of variation	Sum of squares	Degrees of freedom	Mean square
Among journeys			
Overall journey	9,684.00	4	2,421.00
Runs × journey	18,650.07	12	1,554.17
By pot (9 + 12)	9,582.59	16	598.91
Among cylinders			
Overall cylinder	9,132.87	2	4,566.44
Runs × cylinder	11,532.73	6	1,922.12
By pot (10 + 14)	1,466.93	8	183.37
Between pots			
Pots	5,644.41	1	5,644.41
By runs	4,455.16	3	1,485.05
Among runs	13,679.89	3	4,559.96
Residual			
(8 + 11 + 13 + 15)	18,398.14	64	287.47
Total	102,226.79	119	

The accompanying table is equivalent to the table set down by Tippett. Analysis of the mean squares may now be carried out, and the reader is referred to Tippett (42, *a*) for further discussion and final conclusions.

2.19 The *L* tests. It has been noted that a test of the homogeneity of variances (L_1) should precede the F test if the F test is to be construed as a test of the homogeneity of means. The L_1 test was used in the first chapter; we shall now illustrate this test in greater detail.

In addition a new preliminary test known as the L_0 test will be illustrated. The three tests L_0, L_1, and F are most informative when used together, as will be shown presently.

Rider (34, *a*) gives the following Western Electric Company data on the breaking strength in pounds tension of cement briquettes.

Batches									
1	2	3	4	5	6	7	8	9	10
518	508	554	555	536	544	578	530	590	542
560	574	598	567	492	502	532	564	554	556
538	528	579	550	528	548	562	536	530	590
510	534	538	535	572	562	524	540	572	546
544	538	544	540	506	534	548	530	525	522

There are 50 observations, divided among 10 samples, each sample having 5 observations. The notation will be $k = 10$, $n = 5$, and $N = 50$. The questions which the L_0, L_1, and F tests can answer are:

(1) Could these 10 samples belong to normal populations having the same mean and the same variance? (L_0 test.)

(2) Could these 10 samples belong to normal populations of the same variance, no stipulations being made as to the mean? (L_1 test.)

(3) Could these 10 samples belong to normal populations whose means are appreciably the same and whose variances are assumed the same? (F test.)

The functions devised by Neyman and Pearson (30) are respectively:

[5] $$L_0 = \left(\frac{s_1^2 \cdot s_2^2 \cdots s_k^2}{s_0^2 \cdot s_0^2 \cdots s_0^2}\right)^{\frac{1}{k}}$$

[6] $$L_1 = \left(\frac{s_1^2 \cdot s_2^2 \cdots s_k^2}{s_a^2 \cdot s_a^2 \cdots s_a^2}\right)^{\frac{1}{k}}$$

where

$$ns_1^2 = \sum_1^n (X - \bar{X}_1)^2, \quad ns_2^2 = \sum_1^n (X - \bar{X}_2)^2, \ \ldots.$$

$$Ns_0^2 = \sum_1^{nk} (X - \bar{X})^2, \quad Ns_a^2 = n\sum_1^k s_i^2 \quad \text{and} \quad N = nk$$

s_i^2 are the within-sample variances, s_0^2 is the variance based on the deviation of all N observations about their mean $\bar{X}$, and s_a^2 is the mean of the within-sample variances. $\bar{X}_i$ are the sample means. For theoretical reasons, only the case of equal sample (batch) sizes

$$n_1 = n_2 = \cdots = n$$

can be considered.

In both the L_0 and L_1 tests, if the hypothesis that the data do belong to the specified normal population is true, the value of L tends to unity, although the occurrence of unity will be a highly unlikely event even if the hypothesis is true, for L is subject to sampling error. The less the data support the hypothesis, the nearer for given n will the value of the corresponding L come to zero.

The distributions of L_0 and L_1 have been approximated by Neyman and Pearson (30) and tables have been prepared by Mahalanobis (26) (Tables IX and X) and by Nayer (28).

In computing L_0 and L_1, we introduce the geometric mean s_g^2 of the within-sample variances s_i^2.

$$s_g^2 = (s_1^2 \cdot s_2^2 \cdots s_k^2)^{1/k}$$

$$\log s_g^2 = \frac{1}{k} (\log s_1^2 + \log s_2^2 + \cdots + \log s_k^2)$$

Then

$$\log L_0 = \log s_g^2 - \log s_0^2$$

$$\log L_1 = \log s_g^2 - \log s_a^2$$

SAMPLE	WITHIN-SAMPLE VARIANCE s_i^2
1	324.80
2	459.84
3	509.44
4	127.44
5	754.56
6	404.80
7	384.96
8	158.40
9	607.36
10	498.56

$$s_0^2 = 528.96$$

$$s_a^2 = 423.02$$

$$s_g^2 = 374.75$$

from which

$$L_0 = 0.708$$

$$L_1 = 0.886$$

For $n = 5$ and $k = 10$, the 5 per cent level of L_0 (Table IX) is 0.4857. We have $L_0 = 0.708$. The 10 samples are homogeneous with respect to mean and variance.

As the test indicates that the samples came from the same normal population (i.e., from normal populations of the same mean and the same variance) the L_1 and F tests must both fail (i.e., show no significance) for they test the *nature* of any non-homogeneity disclosed by the L_0 test. From Table X, the 5 per cent level of L_1 is 0.6318, indicating that the L_1 hypothesis is upheld; finally, the following analysis of variance shows that the samples are homogeneous in their means, i.e., the null F hypothesis is upheld.

Source of variation	Sum of squares	Degrees of freedom	Mean square
Among batches	5,297.22	9	588 58
Within batches	21,150 80	40	528 77
Total	26,448.02	49	——

The following statistics have been computed from data and calculations given by Dudding and Baker (10). The original data deal with the breaking strain of glass tubing. Each sample contains eight observations.

Sample	Mean	Within-sample variance s_i^2
1	1010	4,025
2	1100	38,013
3	1020	38,488
4	1100	6,700
5	1070	34,300
6	1180	15,475
7	1030	12,375
8	1180	43,100
9	1040	14,775
10	1200	11,238
11	970	38,088
12	1050	50,825
13	840	58,675
14	970	32,588
15	1060	13,413
16	1130	6,838

We find

$$s_0^2 = \frac{4,364,878}{128} = 34,101 \quad \text{(see following analysis of variance)}$$

$$s_a^2 = 26,182$$

$$\log s_g^2 = 4.2999$$

from which

$$L_0 = 0.58$$

$$L_1 = 0.76$$

The 5 per cent levels are, approximately

$$L_0 = 0.62$$

$$L_1 = 0.73$$

The L_0 hypothesis is not supported, i.e., the samples differ significantly in their means and/or variances. The result of the L_1 test indicates that the samples do not differ significantly in their variances; hence they must differ in their means, and an analysis of variance should support this expectation.

Source of variation	Sum of squares	Degrees of freedom	Mean square
Among samples	1,013,550	15	67,570
Within samples (error)	3,351,328	112	29,923
Total	4,364,878	127	......

From the above table

$$F = 2.26$$

which for 15 and 112 degrees of freedom is significant.

A final example is based on data given by Campbell and Lovell (6) on octane ratings of motor fuels. The fuels, which are of known composition, are rated blind eight successive times in eight different makes of car. The results are shown in the two following tables.

63.3 OCTANE NUMBER FUEL

		Rating number							
		1	2	3	4	5	6	7	8
Car	A	63.8	65.7	65.7	65.7	63.8	67.6	67.6	65.8
	B	68.8	55.0	56.3	61.2	61.2	61.2	61.2	61.2
	C	60.0	60.0	60.0	63.8	56.3	67.6	67.6	72.8
	D	58.7	62.2	62.2	63.8	62.2	62.2	62.2	62.2
	E	60.0	63.8	63.8	63.8	60.0	60.0	63.8	63.8
	F	60.2	58.2	63.8	63.8	63.8	63.8	63.8	63.8
	G	63.8	62.5	64.8	64.8	65.8	65.8	63.8	65.8
	H	63.8	63.8	64.8	63.8	63.8	63.8	65.5	65.5

75.0 OCTANE NUMBER FUEL

		Rating number							
		1	2	3	4	5	6	7	8
Car	A	75.0	75.0	73.0	73.0	77.0	76.0	76.0	76.0
	B	75.0	75.0	75.0	71.4	75.0	75.0	75.0	77.0
	C	75.0	75.0	75.0	75.0	75.0	75.0	74.3	75.0
	D	75.0	75.0	77.0	77.0	77.0	77.0	77.0	77.0
	E	75.0	75.0	75.0	73.3	75.0	77.0	77.0	74.0
	F	75.1	75.1	75.1	75.1	75.1	75.1	75.1	78.0
	G	73.0	75.0	75.0	75.0	75.0	75.0	75.0	75.0
	H	75.0	77.0	75.0	72.3	77.0	75.0	75.0	77.0

One of the author's discoveries from these data is that at light knock intensity (75.0 octane number) the variation in knock rating appears less (the standard deviation is 1.2) than for the heavier knock intensity (standard deviation is 3.0). To this they add "No particular significance is attached to the variations of the standard deviations from car to car * * * because of the relatively small amount of data available."

Let us examine this opinion. We find L_1 values of 0.48 and 0.65. The 5 per cent level of L_1 for both fuels is about 0.71. Hence for both grades of fuel the variation in s from car to car is statistically significant.

NOTES

2.20 Cross-product terms in the analysis of variance. In setting the total sum of squares equal to the sum of sums of squares associated with specific factors, certain cross-product terms were assumed to be zero. It is easy to show that they are zero. Thus for the breakdown shown on page 70

$$\begin{aligned}\sum(X - \bar{X})^2 &= \sum[(\bar{X}_l - \bar{X}) + (\bar{X}_{rl} - \bar{X}_l) + (X - \bar{X}_{rl})]^2 \\ &= \sum(\bar{X}_l - \bar{X})^2 + \sum(\bar{X}_{rl} - \bar{X}_l)^2 + \sum(X - \bar{X}_{rl})^2 \\ &\quad + 2\sum(\bar{X}_l - \bar{X})(\bar{X}_{rl} - \bar{X}) + 2\sum(\bar{X}_l - \bar{X})(X - \bar{X}_{rl}) \\ &\quad + 2\sum(\bar{X}_{rl} - \bar{X}_l)(X - \bar{X}_{rl})\end{aligned}$$

Omitting the common multiplier 2, the cross-product terms may be written

$$(\bar{X}_l - \bar{X})\sum(\bar{X}_{rl} - \bar{X}) + (\bar{X}_l - \bar{X})\sum(X - \bar{X}_{rl}) + (\bar{X}_{rl} - \bar{X}_l)\sum(X - \bar{X}_{rl})$$

The first term is an abbreviated statement of

$$\begin{aligned} &(\bar{X}_{l_1} - \bar{X})[(\bar{X}_{r_1 l_1} - \bar{X}) + (\bar{X}_{r_2 l_1} - \bar{X}) + \cdots] \\ &+ (\bar{X}_{l_2} - \bar{X})[(\bar{X}_{r_1 l_2} - \bar{X}) + (X_{r_2 l_2} - \bar{X}) + \cdots] \\ &+ \cdots \end{aligned}$$

The expression outside each bracket is constant for all terms within that bracket. The terms within each bracket represent the sum of the deviations of variates around their own mean; hence each bracket is zero.

Similarly, each of the cross-product terms for the breakdown favored on page 73 is zero. We have

$$\begin{aligned} \sum(X - \bar{X})^2 &= \sum[(\bar{X}_c - \bar{X}) + (\bar{X}_l - \bar{X}) + (\bar{X}_{lc} - \bar{X}_l - \bar{X}_c + \bar{X}) + (X - \bar{X}_{lc})]^2 \\ &= \sum(\bar{X}_c - \bar{X})^2 + \sum(\bar{X}_l - \bar{X})^2 + \sum(\bar{X}_{lc} - \bar{X}_l - \bar{X}_c + \bar{X})^2 + \sum(X - \bar{X}_{lc})^2 \\ &+ 2\sum(\bar{X}_c - \bar{X})(\bar{X}_l - \bar{X}) + 2\sum(\bar{X}_c - \bar{X})(\bar{X}_{lc} - \bar{X}_l - \bar{X}_c + \bar{X}) \\ &+ 2\sum(\bar{X}_c - \bar{X})(X - \bar{X}_{lc}) + 2\sum(X_l - \bar{X})(\bar{X}_{lc} - \bar{X}_l - \bar{X}_c + \bar{X}) \\ &+ 2\sum(\bar{X}_l - \bar{X})(X - \bar{X}_{lc}) + 2\sum(\bar{X}_{lc} - \bar{X}_l - \bar{X}_c + \bar{X})(X - \bar{X}_{lc}) \end{aligned}$$

By the arguments used in the preceding example the third, fifth, and sixth cross-product terms are zero. Omitting the factor 2, the remaining cross-products may be written

$$\begin{aligned} \cancel{\sum(\bar{X}_c - \bar{X})(\bar{X}_l - \bar{X})} + \sum(\bar{X}_c - \bar{X})(\bar{X}_{lc} - \bar{X}_c) - \sum(\bar{X}_c - \bar{X})(\bar{X}_l - \bar{X}) \\ + \sum(\bar{X}_l - \bar{X})(\bar{X}_{lc} - \bar{X}_l) - \cancel{\sum(\bar{X}_l - \bar{X})(\bar{X}_c - \bar{X})} \end{aligned}$$

As in the previous example the second and fourth terms are zero. The other term may be expanded with one term in parentheses outside the summation. The term within the summation is the sum of the deviation of variates around their means and hence is zero.

2.21 Distribution of *F*. The argument underlying the comparison of an index of variation due to a suspected " cause " with a similar index associated with unknown ($\equiv$ chance) causes is the following:

If an indefinitely large number of random samples are drawn from a known population, a sample statistic (such as the mean or variance) will have a continuous distribution curve which can often be exactly determined by mathematical procedure rather than be approximated by any amount of experimental sampling. For example, if an indefinitely large number of random samples each of n independent observations are drawn from a normal population of variance σ^2, and if for each sample the statistic

$$\chi^2 = \frac{ns^2}{\sigma^2}$$

is computed, the distribution of χ^2, with $n - 1$ degrees of freedom, is

$$C\chi^{2\,(n-3)/2}\, e^{-\chi^2/2}\, d\chi^2$$

where C is a constant.

Similarly, if we have two unbiased estimates $\hat{\sigma}_x^2$ and $\hat{\sigma}_y^2$ of the variance σ^2 of a normal population where

$$\hat{\sigma}_x^2 = \frac{\sum(X - \bar{X})^2}{f_x}$$

$$\hat{\sigma}_y^2 = \frac{\sum(Y - \bar{Y})^2}{f_y}$$

f_x and f_y being the number of degrees of freedom on which each estimate is based, the ordinate of the distribution of the ratio

$$F = \frac{\hat{\sigma}_x^2}{\hat{\sigma}_y^2}$$

is found to be

$$\psi(f_1, f_2) \cdot \frac{F^{(f_1-2)/2}}{(f_1F + f_2)^{(f_1+f_2)/2}}$$

ψ being known. This distribution is found in the following way: The distribution of $\hat{\sigma}_x^2$ (for a given value of $\hat{\sigma}_y^2$) is known. Similarly for $\hat{\sigma}_y^2$. The distribution of their ratio F is found by multiplying the distribution function of $\hat{\sigma}_x^2$ (for a given $\hat{\sigma}_y^2$) by the distribution function of $\hat{\sigma}_y^2$ and integrating the product over all values of $\hat{\sigma}_y^2$ (0 to ∞). Table VIII gives values of F, for various f_1 and f_2, beyond which lies 5 per cent (and 1 per cent) of the area under the curve of F, i.e., values of F satisfying the equation

$$\begin{matrix} 0.05 \\ \text{or} \\ 0.01 \end{matrix} = \int_F^\infty \psi(f_1, f_2) \cdot \frac{F^{(f_1-2)/2}}{(f_1F + f_2)^{(f_1+f_2)/2}}\, dF$$

To summarize: we compute the ratio F from the data. We then determine the probability in random sampling from a normal population that the computed ratio would be exceeded. If this probability is small (0.05 or 0.01) we conclude that the mean square in the numerator of F is significantly greater than the true estimate of σ^2 furnished by the denominator; the mean squares could not have arisen from the same normal population — variation associated with that cause is statistically significant.

2.22 Estimating σ^2. We now outline a proof that the mean squares given in the last column of each analysis of variance table are unbiased estimates of the population variance σ^2.

The method has been used by Irwin (22). Let N observations X_{ij} be divided among R rows and C columns, $RC = N$.

		Column		
		1	$2 \cdots$	C
	1	X_{11}	$X_{12} \cdots$	X_{1c}
	2	X_{21}	$X_{22} \cdots$	X_{2c}
Row				
	$\cdot$	$\cdot$	$\cdot$	$\cdot$
	$\cdot$	$\cdot$	$\cdot$	$\cdot$
	$\cdot$	$\cdot$	$\cdot$	$\cdot$
	R	X_{r1}	$X_{r2} \cdots$	X_{rc}

Source of variation	Sum of squares	Degrees of freedom	Mean square
Rows	$\sum(\bar{X}_r - \bar{X})^2$	$R - 1$	$\dfrac{\sum(\bar{X}_r - \bar{X})^2}{R - 1}$
Columns	$\sum(\bar{X}_c - \bar{X})^2$	$C - 1$	$\dfrac{\sum(\bar{X}_c - \bar{X})^2}{C - 1}$
Interaction	$\sum(X - \bar{X}_r - \bar{X}_c + \bar{X})^2$	$(R - 1)(C - 1)$	$\dfrac{\sum(X - \bar{X}_r - \bar{X}_c + \bar{X})^2}{(R - 1)(C - 1)}$
Total	$\sum(X - \bar{X})^2$	$RC - 1$	$\dfrac{\sum(X - \bar{X})^2}{RC - 1}$

It has already been shown that

$$\frac{\sum(X - \bar{X})^2}{RC - 1}$$

is an unbiased estimate of σ^2, based on $RC - 1$ degrees of freedom. Now for the variance due to rows; write the expected value (mean value over all samples) of $(\bar{X}_r - \bar{X})^2$ as follows:

[7] $$E[(\bar{X}_r - \bar{X})^2] = E\left[\frac{1}{C}\sum^{C}\left\{(X - \bar{X}') - (\bar{X} - \bar{X}')\right\}\right]^2$$

Expanding the right-hand side and writing down the expected values of the resulting three functions, for example,

$$E[\sum(X - \bar{X}')^2] = \sigma^2$$

we find

$$E[\sum(\bar{X}_r - \bar{X})^2] = \sigma^2(r - 1)$$

or

$$\frac{\sum(\bar{X}_r - \bar{X})^2}{R - 1}$$

is an unbiased estimate of σ^2 based on $R - 1$ degrees of freedom.

X is normally distributed, hence the mean $\bar{X}_r$ is also normally distributed; the distribution of

$$\frac{\sum(X_r - \bar{X})^2}{R - 1}$$

is essentially that of s^2.

Similarly,

$$\frac{\sum(\bar{X}_c - \bar{X})^2}{C - 1}$$

is an unbiased estimate of σ^2 based on $c - 1$ degrees of freedom.

In the case of the sum of squares

$$\sum(X - \bar{X}_r - \bar{X}_c + \bar{X})^2$$

we have

$$E(X - \bar{X}_r - \bar{X}_c + \bar{X})^2 = E[(X - \bar{X}') - (\bar{X}_r - \bar{X}') - (\bar{X}_c - \bar{X}') + (\bar{X} - \bar{X}')]^2$$

After expansion and use of [7], we find

$$E[\sum(X - \bar{X}_r - \bar{X}_c + \bar{X})^2] = (R - 1)(C - 1)\sigma^2$$

or

$$\frac{\sum(X - X_r - X_c + \bar{X})^2}{(R - 1)(C - 1)}$$

is an unbiased estimate of σ^2 based on $(R - 1)(C - 1)$ degrees of freedom.

CHAPTER III

RELATIONSHIP AMONG VARIABLES

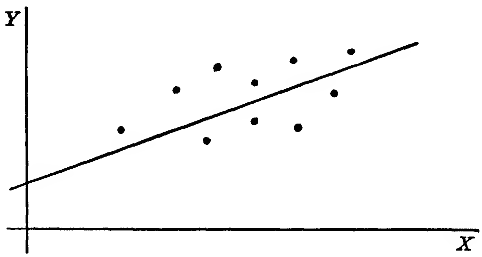
3.1 Introduction. In several sciences, for example, physics, relationships among variables are often stated in exact functional form. Thus the relationship between time and distance for an object falling in a vacuum is written simply as $s = \frac{1}{2}gt^2$, and it is implied that s is exactly determinable from t.

Without questioning the validity of this practice in physics, it is apparent that it is not valid in industrial research. Both the nature of industrial experimenting and the impracticability of duplicating the relatively controlled conditions of physics laboratories bring about this result. The hardness and tensile strength of one aluminum casting may be respectively X and Y, while for a second apparently identical specimen we find X and $1.2Y$. Specification of Y from X is subject to error, i.e., from knowledge of X we can estimate only the average (expected) value of Y, not the value which is actually observed.

3.2 Types of relationships. Such relationships among variables can be classified. If there are two variables whose relationship is described by a straight line, the term linear regression is used to describe the relationship. If the relationship is parabolic, say $Y = aX + bX^2$ where a and b are constants, the term curvilinear regression is used. For a relationship among more than two variables, such as $Y = aU + bV$, where a and b are constants, the term multiple regression is used. The relationship among any k of n related variables, the remaining $n - k$ variables being in the simplest case, held constant, is described as partial correlation or partial regression. We shall presently illustrate simple linear and simple multiple regression.

3.3 Uses of regression analysis. Regression analysis is useful wherever hypotheses dealing with relationships are examined. To give a few examples: in agriculture the relationship between crop size and tree injury has been studied; in medicine, studies of the relationship between vitamin potency and weight gain have successfully used regression analysis; in economics, and other social sciences where perhaps serious questions regarding the validity of the technique can be raised, the use of regression analysis has been extensive and there seems to be hardly any group of variables to which simple, multiple, and partial cor-

relation and regression analysis has not been applied. In industrial research regression analysis can be used in the search for inexpensive methods of testing as replacements for more expensive methods. We shall give several examples of this usage.



3.4 General procedure. Let it be presumed that two variables are linearly related. We construct the best fitting straight line.

$$Y_r = a + bX$$

The total variation of say the Y-value of an observation about its mean $\bar{Y}$ can be represented by

$$(Y - \bar{Y})$$

where Y represents an individual observation. This may be divided into two parts: first, a part explained by the relationship of Y to X, and second, any remainder.

Consider the first part. If Y and X are unrelated, the expected or best value of Y, given X, is $\bar{Y}$. If Y and X are related, the best estimate of Y, given X is Y_r, the ordinate of the regression line at X. The greater the relationship between the two variables Y and X, the greater the superiority of Y_r over $\bar{Y}$ as an estimate of Y, for given X. Hence the term

$$(Y_r - \bar{Y})$$

represents that part of the total variation allocable to regression.

The remainder consists of variation unallocable to regression, and inasmuch as there are no other specific factors to which such variation can be allocated, this term is of purely residual character. It consists of the variation of observations about the regression line and is given by

$$(Y - Y_r)$$

In measuring these three classes of variation, it might be supposed that we use

$$\sum(Y - \bar{Y}) = \sum(Y_r - \bar{Y}) + \sum(Y - Y_r)$$

the summation extending over all n values of Y or Y_r. This equation is valid but not useful, for $\sum(Y - \bar{Y})$ is always zero regardless of the

amount of the variation of Y_i about $\bar{Y}$. If, however, variation is measured by the squares of deviations we shall have

$$\sum(Y - \bar{Y})^2 = \sum(Y_r - \bar{Y})^2 + \sum(Y - Y_r)^2$$

as will presently be proved.

To obtain three estimates of the population variance from these sums of squares, it is necessary to introduce, as before, the idea of degrees of freedom. There are n values of Y in $\sum(Y - \bar{Y})^2$ less the grand mean which is computed from Y, or $n - 1$ degrees of freedom. There are n values of Y in $\sum(Y - Y_r)^2$ less the computed constants of Y_r (a and b), or $n - 2$ degrees of freedom. By subtraction there remains only one degree of freedom for the linear regression estimate based on $\sum(Y_r - \bar{Y})^2$, which is as it should be for the two constants of the regression line (a and b) are restricted by $\bar{Y}$.

Source of variation	Sum of squares	Degrees of freedom	Mean square
Linear regression	$\sum(Y_r - \bar{Y})^2$	1	$\hat{\sigma}_1^2 = \dfrac{\sum(Y_r - \bar{Y})^2}{1}$
Residual	$\sum(Y - Y_r)^2$	$n - 2$	$\hat{\sigma}_2^2 = \dfrac{\sum(Y - Y_r)^2}{n - 2}$
Total	$\sum(Y - \bar{Y})^2$	$n - 1$	

The test of significance is again the F test. If there is no real linear relationship between Y and X, the two mean squares should be the same. If, however, the value of F in

$$F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}$$

is sufficiently large, the regression is real.

Thus if we make many drawings (of size n) from a bowl of chips, each chip being marked with two numbers, one being X and one Y, the distribution of X and Y being bivariate normal, with the overall correlation between X and Y in the bowl being zero, and if from each drawing we construct $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$, the distribution of their ratio F for all drawings of size n could be determined. It is this distribution whose values are shown in Table VIII. For example, for $n = 20$ only 5 per cent of such drawings will yield values of F exceeding 4.38 (1 and 19 degrees of freedom).

The results obtained in an ordinary industrial experiment are judged by the probabilities given in Table VIII. Thus if we find $F = 10$ for 1 and 19 degrees of freedom, we conclude that there is less than 1 chance in 100 that $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are estimates of the variance of the same homogeneous normal population. The hypothesis is rejected and we shall conclude that the populations are not identical and that the regression is real.

3.5 Fitting the regression line. The criterion we shall use in fitting the regression line $Y_r = a + bX$ is the following: if one of the two variables X and Y (say Y) is subject to error while the other is not, the sum of the squares of vertical distances from the Y's to the corresponding Y_r, i.e., $\sum(Y - Y_r)^2$, is to be a minimum. This requirement yields two equations which are solved for a and b. This and other properties of a regression line fitted to observations by the method of least squares will be developed later.

3.6 Examples of linear regression. Brenner (4) gives the following data on the thickness in hundred-thousandths of an inch of non-magnetic coatings of galvanized zinc on 11 pieces of iron and steel:

Thickness as measured by standard destructive stripping method	Thickness as measured by non-destructive magnetic method
Y	X
116	105
132	120
104	85
139	121
114	115
129	127
720	630
174	155
312	250
338	310
465	443

Measurement of thickness by stripping is accurate but the tests are destructive and costly. The magnetic method is less costly. Do the data support the belief that we may measure X and use Y_r as an estimate of Y where

$$Y_r = a + bX$$

a and b being constants?

The following is a résumé of the procedure described in 3.4. The significance of a straight line

$$Y_r = a + bX$$

fitted to data by the method of least squares, may be tested by determining the probability, in random sampling from a normal population, that the computed value of

$$F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}$$

would be exceeded, where

$$\hat{\sigma}_1^2 = \frac{\sum (Y_r - \bar{Y})^2}{1}$$

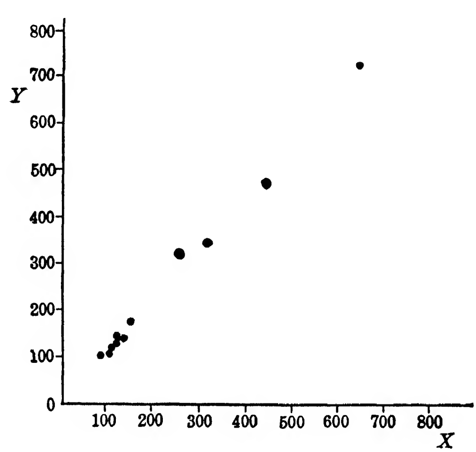
is the mean square attributable to regression, with 1 degree of freedom, and

$$\hat{\sigma}_2^2 = \frac{\sum (Y - Y_r)^2}{n - 2}$$

the residual or chance mean square, not accounted for by regression, with $n - 2$ degrees of freedom.

$\bar{Y}$ is the mean of Y and n is the number of pairs of observations.

The Y values appear not to have come from a normal population but experimental evidence on this point indicates that the F test can probably be used. The points are plotted in the following diagram.



Using the method of least squares to determine a and b,

$$a = \frac{\sum X^2 \sum Y - \sum X \sum XY}{n\sum X^2 - (\sum X)^2}$$

$$b = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2}$$

From the data,

$$\sum Y = 2{,}743$$

$$\sum X = 2{,}461$$

$$\sum XY = 952{,}517$$

$$\sum Y^2 = 1{,}067{,}143$$

$$\sum X^2 = 852{,}419$$

$$n = 11$$

from which

$$a = -1.7948$$

$$b = 1.1226$$

Using these values of a and b, the predicted values of thickness calculated from a knowledge of the magnetic readings X are shown in the following table.

Predicted values of thickness $Y_r = a + bX$	True values of thickness Y
116.08	116
132.92	132
93.64	104
134.04	139
127.31	114
140.78	129
705.43	720
172.21	174
278.86	312
346.21	338
495.52	465

Is the discrepancy between these pairs of values small, from a statistical point of view? The answer is found, i.e., the adequacy of a linear

equation is determined, by comparing that part of the total variability for which the line can account with remaining unaccountable (chance) variability.

Source of variation	Sum of squares	Degrees of freedom	Mean square
Regression	$\sum(Y_r - \bar{Y})^2$	1	$\sum(Y_r - \bar{Y})^2/1$
Residual	$\sum(Y - Y_r)^2$	$n - 2$	$\sum(Y - Y_r)^2/n - 2$
Total	$\sum(Y - \bar{Y})^2$	$n - 1$	...

We have

$$\sum(Y - \bar{Y})^2 = \sum Y^2 - \frac{(\sum Y)^2}{n} = 1{,}067{,}143 - \frac{(2{,}743)^2}{11}$$

$$= 383{,}138.54$$

$$\sum(Y_r - \bar{Y})^2 = \sum Y_r^2 - \frac{(\sum Y)^2}{n} = 1{,}064{,}356.41 - \frac{(2{,}743)^2}{11}$$

$$= 380{,}377.41$$

from which, by subtraction,

$$\sum(Y - Y_r)^2 = 2{,}761.13$$

Source of variation	Sum of squares	Degrees of freedom	Mean square
Regression	380,377.41	1	380,377.41
Residual	2,761.13	9	306.79
Total	383,138.54	10	...

If the regression line is inadequate, the mean square due to regression will not be significantly larger than the residual or chance mean square. In our case, the ratio is

$$F = \frac{380{,}377.41}{306.79} = 1240$$

Table VIII gives the values of F which, with one and nine degrees of freedom, are necessary in order to establish regression as (1) significant

(2) highly significant. F need be only 10.56 for the regression to be highly significant. Since we have $F = 1240$, our estimates of thickness from magnetic measurements by use of the equation

$$Y_r = -1.7948 + 1.1226X$$

are statistically sound. The practical man must now decide if the discrepancies between Y and Y_r are sufficiently small, from the point of view of the use to which the product is put.

Jennett and Dudding (24) give the following 11 observations on life tests of electric light bulbs and tests on filament wire.

Life of bulb in hours Y	Quality test on filament X
1,605	276
1,120	293
1,320	288
1,225	315
1,055	305
1,390	315
1,385	306
1,700	286
2,070	289
1,395	296
1,105	335

A life test required about 1000 hours and cost about \$5 per bulb. The wire test is quickly performed and is lower in cost. If only wire tests are made, can life be estimated from

$$Y_r = a + bX$$

As before, we have

$$\sum Y = 15{,}370$$

$$\sum X = 3{,}304$$

$$\sum XY = 4{,}588{,}135$$

$$\sum X^2 = 995{,}238$$

$$n = 11$$

from which $a = 4{,}410.296$ and $b = -10.031$.

Source of variation	Sum of squares	Degrees of freedom	Mean square
Regression	285,429.57	1	285,427.83
Residual	617,238.62	9	68,582.01
Total	902,668.19	10	. . .

Finally

$$F = 4.1$$

From Table VIII, with one and nine degrees of freedom F need be as high as 10.56 in order for the linear regression to be highly significant (1 per cent level) or 5.12 for significance at the customary level (5 per cent). $F = 4.1$ does not meet these requirements. Linear regression does not account for a sufficient part of the total variability; it is not adequate for the purpose of prediction. The residual variability about the regression line

$$\frac{\sum(Y - Y_r)^2}{n - 2}$$

is too large. Values of Y_r are calculated from

$$Y_r = 4{,}410.296 - 10.031X$$

and the following table shows Y_r and Y:

PREDICTED VALUES OF LIFE Y_r	ACTUAL LIFE TESTS Y
1,641.7	1,605
1,471.2	1,120
1,521.4	1,320
1,250.5	1,225
1,350.8	1,055
1,250.5	1,390
1,340.8	1,385
1,541.4	1,700
1,511.3	2,070
1,441.1	1,395
1,049.9	1,105

The correspondence of Y_r and Y is not sufficiently high.

It is impossible to state flatly that a sample of 11 is too small but samples perhaps of 40 or more observations are desirable if tentative conclusions of industrial importance are to be drawn from the results.

As an example from chemical research work, consider the following data given by Thomsen (41); we wish to predict titer values from the iodine values of fatty acids.

Y Titer value (minus 40)	X Iodine value (minus 40)	Y Titer value (minus 40)	X Iodine value (minus 40)
2.5	7.2	2.0	14.2
2.0	10.3	1.8	10.3
3.0	7.9	5.4	0.5
3.2	8.2	3.1	9.4
2.1	13.7	4.8	1.4
4.8	1.9	3.7	5.9
2.1	13.0	0.1	16.4
1.5	17.5	1.3	17.3
4.8	0.3	4.8	0.8
3.8	7.3	1.3	16.4
2.2	12.1	0.0	12.2
0.4	18.5	5.0	2.5
6.6	−1.3	2.4	13.4
4.3	1.9	5.7	0.3
2.0	13.8	3.9	4.1
1.2	15.1	2.1	11.4
2.3	8.8	0.3	21.7
3.2	6.3	3.5	5.7
1.2	15.4	4.3	3.5
5.3	−0.2	4.5	3.8
4.0	0.7	2.2	11.1
2.9	9.1	3.4	7.6
3.0	8.8	2.3	2.8
2.9	9.1	4.9	1.9
3.1	8.8	1.7	17.8

We have

$$\sum Y = 148.9$$

$$\sum X = 426.6$$

$$\sum XY = 853.37$$

$$\sum Y^2 = 561.77$$

$$\sum X^2 = 5{,}387.08$$

$$n = 50$$

The accompanying table shows the analysis of variance.

Source of variation	Sum of squares	Degrees of freedom	Mean square
Regression	99.5407	1	99.5407
Residual	18.8051	48	0.3918
Total	118.3458	49	...

Forty has been subtracted from the observed values of both X and Y; this facilitates computation, for the smaller numbers are easier to handle and it does not affect the numerical analysis.

We have $F = 254.08$. The 1 per cent value of F is 7.19; the regression (which is negative, i.e., high values of one variable are generally associated with low values of the other variable and vice versa) is highly significant.

3.7 Linear regression in grouped data. In the previous examples we had at most 50 observations, so there was no reason to group the observations. In the present example there were originally 440 observations; they have been grouped into classes in the table below. In such a case the analysis is slightly more complex.

It is here assumed that the variances of all eight columns are equivalent, within the limits of chance variation. This assumption, which may be checked by the L_1 test, must be met for the F test to be valid.

The British Cotton Industry Research Association (5) gives the following data on the frequency of warp breaks in weaving, classified according to values of an important influence, namely, relative humidity.

Relative humidity (X)

		68–	70–	72–	74–	76–	78–	80–	82–	Total
Warp breaks (Y)	0.0–				2	5	5	1		13
	0.8–		1	7	13	28	28	11	4	92
	1.6–	2	6	16	27	44	35	9	1	140
	2.4–	1	5	24	24	27	17	3	2	103
	3.2–		2	16	6	15	9	2	1	51
	4.0–	2	1	4	7	7	5	1		27
	4.8–		1	2	2	3	2			10
	5.6–			1			2			3
	6.4–						1			1
	Total	5	16	70	81	129	104	27	8	440

Would the regression line

$$Y_r = a + bX$$

enable us effectively to predict warp breakage, or is the residual variability too great?

We shall first proceed as before.

$$\sum Y = 1{,}063.2$$

$$\sum X = 33{,}446$$

$$\sum XY = 80{,}524.4$$

$$\sum Y^2 = 3{,}110.4$$

$$\sum X^2 = 2{,}545{,}724$$

$$n = 440$$

from which

$$a = 9.028$$

$$b = -0.087$$

The analysis of variance is given in the following table.

Source of variation	Sum of squares	Degrees of freedom	Mean square
Regression	25.51	1	25.51
Residual	515.81	438	1.18
Total	541.32	439	...

An F test indicates significant regression. But the sum of squares not due to regression, $\sum(Y - Y_r)^2$, consists of two parts — the sum of squares of the deviations of the column means $\bar{Y}_c$ about the corresponding Y_r, i.e., $\sum(\bar{Y}_c - Y_r)^2$ plus the variability within columns $\sum(Y - \bar{Y}_c)^2$, which is more truly the chance or unallocable variability.

The total residual degrees of freedom were $n - 2$. The unallocable part $\sum(Y - \bar{Y}_c)^2$ uses k means computed from the data: hence the unallocable mean square is $\dfrac{\sum(Y - \bar{Y}_c)^2}{n - k}$. By subtraction, the deviation-from-regression mean square has $k - 2$ degrees of freedom.

Thus for linear regressions calculated from grouped data, the following subsidiary variance analysis may be useful.

Source of variation	Sum of squares	Degrees of freedom	Mean square
Deviation of means from regression	$\sum(\bar{Y}_c - Y_r)^2$	$k - 2$	$\dfrac{\sum(\bar{Y}_c - Y_r)^2}{k - 2}$
Unallocable part of residual (chance)	$\sum(Y - \bar{Y}_c)^2$	$n - k$	$\dfrac{\sum(Y - \bar{Y}_c)^2}{n - k}$
Residual	$\sum(Y - Y_r)^2$	$n - 2$	...

We may compute the "chance" sum of squares and then determine the deviation-from-regression sum of squares by subtraction.

It should be observed that all summations extend over the entire data. Hence, each mean must be counted as many times as there are observations from which that mean was computed. This was discussed in Section 2.10.

Source of variation	Sum of squares	Degrees of freedom	Mean square
Deviation of means from regression	3.37	6	0.57
Unallocable part of residual (chance)	512.44	432	1.19
Residual	515.81	438	...

The earlier F test

$$F = \frac{25.51}{1.18}$$

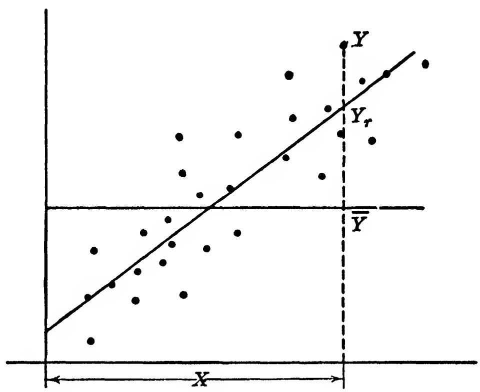
for 1 and 438 degrees of freedom showed the regression to be highly significant. The present value of F is the ratio of 25.51 to 1.19, for 1 and 432 degrees of freedom and the conclusion originally reached is shown to be valid. In the present example the subsidiary analysis of variance was uninformative. If, however, the original test showed the regression not to be significant, any reduction of the residual mean square by elimination of the (possibly) allocable element (deviation from regression) might show the regression to be significant, and the latter inference would be the proper one.

NOTES

3.8 Breakdown of $\sum(Y - \bar{Y})^2$. The accompanying graph may be helpful. It is apparent from the graph that an observation Y may always be written

$$Y = \bar{Y} + (Y_r - \bar{Y}) + (Y - Y_r)$$

the first term to the right of the equality sign ($\bar{Y}$) being common to all observations, the second term ($Y_r - \bar{Y}$) representing that part of the value of the observation Y attributable to regression, and the last term ($Y - Y_r$) repre-



senting the unallocable (and therefore dealt with as chance) variability about the line of regression. The total deviation ($Y - \bar{Y}$) thus consists of two parts, regression ($Y_r - \bar{Y}$) and residual ($Y - Y_r$).

$$(Y - \bar{Y}) = (Y_r - \bar{Y}) + (Y - Y_r)$$

Summing this linear expression merely leads to zero = zero. We must show that

[1] $$\sum(Y - \bar{Y})^2 = \sum(Y_r - \bar{Y})^2 + \sum(Y - Y_r)^2$$

We have

[2] $$\sum(Y - \bar{Y})^2 = \sum[(Y - Y_r) + (Y_r - \bar{Y})]^2$$
$$= \sum(Y - Y_r)^2 + \sum(Y_r - \bar{Y})^2 + 2\sum(Y - Y_r)(Y_r - \bar{Y})$$

We now show that if the regression line

$$Y_r = a + bX$$

is fitted to the data by the method of least squares, the cross product term of [2] is zero and [1] is valid.

3.9 The method of least squares. In the method of least squares, the function $Y_r = a + bX$ is fitted to the data so that

$$\sum(Y - Y_r)^2$$

which will be designated by φ, is minimum. The necessary conditions are

$$\frac{\partial}{\partial a}\sum(Y - a - bX)^2 = 0$$

$$\frac{\partial}{\partial b}\sum(Y - a - bX)^2 = 0$$

Differentiation yields

$$[3] \qquad \begin{aligned} -2\sum(Y - a - bX) &= 0 \\ -2\sum(Y - a - bX)X &= 0 \end{aligned}$$

It is apparent that $\partial^2\varphi/\partial a^2$ and $\partial^2\varphi/\partial b^2$ are always positive, i.e., [3] are conditions for minimum φ. [3] may be written

$$\sum Y = na + b\sum X$$

$$\sum YX = a\sum X + b\sum X^2$$

These " normal " equations may be solved simultaneously for a and b, the Y-axis intercept and the slope of the regression line.

We return to the cross-product term of [2], which may be written

$$\begin{aligned} &\sum(Y - Y_r)Y_r - \bar{Y}\sum(Y - Y_r) \\ &= \sum(Y - Y_r)(a + bX) - \bar{Y}\sum(Y - Y_r) \\ [4] \qquad &= a\sum(Y - Y_r) + b\sum(Y - Y_r)X - \bar{Y}\sum(Y - Y_r) \end{aligned}$$

From [3] it is clear that [4] is 0. Hence [1] is valid.

3.10 Curvilinear regression. If a curvilinear regression function, say the parabola

$$Y_r = a + bX + cX^2$$

(with two degrees of freedom) or an m-degree function (with m degrees of freedom) is fitted to the data by the method of least squares, the cross-product term is easily shown to be 0. The normal equations for determining a, b, and c would be

$$\sum Y = na + b\sum X + c\sum X^2$$

$$\sum YX = a\sum X + b\sum X^2 + c\sum X^3$$

$$\sum YX^2 = a\sum X^2 + b\sum X^3 + c\sum X^4$$

The extension to the general case of an mth degree polynomial is obvious.

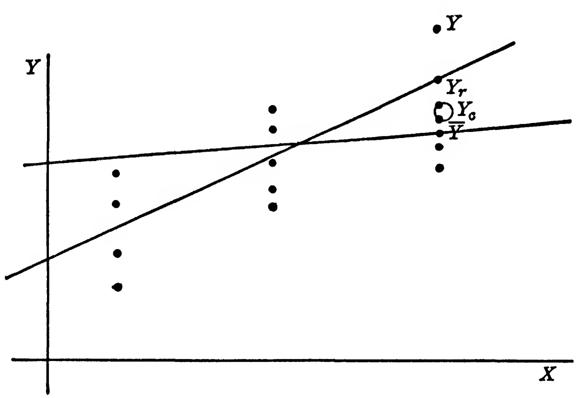
A high degree regression function may fit a set of observations " better " than a low degree function, but the error term may be increased by loss of degrees of freedom. Lower degree functions, such as the straight line or the parabola, are to be preferred for it is best to formalize observed relationships as simply as possible. For it is often impossible — particularly in industrial research — to rationalize (i.e., to explain by recourse to available theory) any but these simpler relationships. Higher-degree regression functions may lead to new theory, but the use of simpler relationships is in keeping with the conservative methodology of experimental science.

3.11 Degrees of freedom. A basic explanation of the allocation of the total number of degrees of freedom is beyond the possibilities of this book, but the following may be helpful. The quantity $\sum(Y - \bar{Y})^2$ summed over n observations has $n - 1$ degrees of freedom, for the mean $\bar{Y}$ has been calculated from the n observations. The regression sum of squares $\sum(Y_r - \bar{Y})^2$ may be written

$$\sum(Y_r - \bar{Y})^2 = \sum(\bar{Y} + b(X - \bar{X}) - \bar{Y})^2$$
$$= b^2\sum(X - \bar{X})^2$$

$X - \bar{X}$ is independent of any correlation between X and Y. Hence variability in $\sum(Y_r - \bar{Y})^2$ depends on b; accordingly only one degree of freedom is allocated to regression. The residual sum of squares $\sum(Y - Y_r)^2$ absorbs the remaining $n - 2$ degrees of freedom.

3.12 Linear regression in grouped data. The accompanying diagram refers to the problem of linear regression in grouped data.



$$Y = \bar{Y} + (Y_r - \bar{Y}) + (Y - Y_r)$$
$$= \bar{Y} + (Y_r - \bar{Y}) + (\bar{Y}_c - Y) + (Y - \bar{Y}_c)$$

To show that

[5] $$\sum(Y - Y_r)^2 = \sum(\bar{Y}_c - Y_r)^2 + \sum(Y - \bar{Y}_c)^2$$

write

$$\sum(Y - Y_r)^2 = \sum[(\bar{Y}_c - Y_r) + (Y - \bar{Y}_c)]^2$$

$$= \sum(\bar{Y}_c - Y_r)^2 + \sum(Y - \bar{Y}_c)^2 + 2\sum(\bar{Y}_c - Y_r)(Y - \bar{Y}_c)$$

But

$$\sum(\bar{Y}_c - Y_r)(Y - \bar{Y}_c)$$

$$= \sum\bar{Y}_c(Y - \bar{Y}_c) - \sum Y_r(Y - \bar{Y}_c)$$

$$= \quad 0 \quad - \quad 0$$

or [5] is valid.

3.13 Computation of sums of squares for linear regression analysis. To compute $\sum(Y_r - \bar{Y})^2$ we may use the expansion

$$\sum(Y_r - \bar{Y})^2 = \sum Y_r^2 - 2\bar{Y}\sum\bar{Y}_r + \sum\bar{Y}^2$$

$$= \sum Y_r^2 - n\bar{Y}^2 = \sum Y_r^2 - \frac{(\sum Y)^2}{n}$$

for, from the normal equations

$$\frac{\sum Y_r}{n} = \bar{Y}$$

It is, however, not necessary to calculate Y_r to find $\sum(Y_r - \bar{Y})^2$ for

$$\sum(Y_r - \bar{Y})^2 = \sum(a + bX - \bar{Y})^2$$

$$= b^2\sum(X - \bar{X})^2$$

$$= \frac{\left[\sum XY - \frac{\sum X \sum Y}{n}\right]^2}{\left[\sum X^2 - \frac{(\sum X)^2}{n}\right]}$$

$$= b\left(\sum XY - \frac{\sum X \sum Y}{n}\right)$$

Convenient expansions for the other sums of squares are

$$\sum(Y - \bar{Y})^2 = \sum Y^2 - n\bar{Y}^2 = \sum Y^2 - \frac{(\sum Y)^2}{n}$$

and

$$\begin{aligned}\sum(Y - Y_r)^2 &= \sum(Y - Y_r)(Y - Y_r)\\ &= \sum(Y - Y_r)Y - a\sum(Y - Y_r) - b\sum(Y - Y_r)X\\ &= \sum(Y - Y_r)Y - \quad 0 \quad - \quad 0\\ &= \sum Y^2 - a\sum Y - b\sum XY\end{aligned}$$

3.14 Regression and prediction (see Eisenhart, 12). The regression line

$$Y_r = a + bX$$

obtained by minimizing

$$\sum(Y - Y_r)^2$$

X being the independent variable, differs from the regression line $X_r = c + dY$ obtained if

$$\sum(X - X_r)^2$$

is minimized. The decision as to which line (or curve) is appropriate (i.e., which of the two variables, X and Y, is to be considered independent) depends not on what we would like to predict, but on which of the two variables, X and Y, is free from error. If we are studying the relationship between quality of output (Y) and time (X), the latter will generally be represented by values (selected in advance of measurement of Y) say, at daily or weekly intervals; such selected values are free from error. Measurements of Y are, however, subject to error and the appropriate regression line would be that which minimized $\sum(Y - Y_r)^2$, that is

$$Y_r = a + bX$$

This is the linear regression of Y on X. It is important to note that in the theory of regression, only the dependent variable (Y in the above example) is required to be normally distributed. The correctness of regression analysis is unimpaired by the fact that the X values are arbitrarily selected, for example, uniformly spaced.

Many problems in industrial research do not lend themselves to a clear-cut decision as to which variable is free from error. In the example on titer and iodine, neither variable appears to have been selected, and both X and Y vary normally and one regression line is apparently as good as the other. A better solution to this problem has been suggested by Wald (44).

Occasionally the variable to be estimated, say Y may not be subject to error whereas X is subject to error. In this case we would first obtain the regression of X on Y

$$X_r = c + dY$$

and determine Y from

$$Y = \frac{X_r - c}{d}$$

3.15 Example of multiple regression. Fulweiler, Stang, and Sweetman (18) give the following data on worn wire rope of nominal diameter ½ to ⅝ inch. Can tensile strength be estimated from a linear relationship connecting U and V?

Tensile strength in thousands of pounds per square inch	Number of broken wires in worst lay	Length of worn surface	Tensile strength in thousands of pounds per square inch	Number of broken wires in worst lay	Length of worn surface
Y	U	V	Y	U	V
174	8	0.14	178	9	0.14
185	0	0.00	185	12	0.14
188	8	0.12	172	0	0.00
160	14	0.11	158	8	0.11
179	0	0.00	162	14	0.11
183	0	0.11	176	0	0.00
191	0	0.09	192	29	0.13
177	0	0.00	198	37	0.13
183	0	0.12	158	0	0.00
186	0	0.13	152	6	0.10
180	0	0.19	136	7	0.08
184	3	0.15	174	0	0.00
175	5	0.13	180	2	0.15
175	0	0.00	172	5	0.15
166	2	0.16	172	0	0.00
170	14	0.15	174	0	0.13
180	0	0.00	162	5	0.13
181	0	0.12	165	0	0.00
201	12	0.15	153	12	0.11
172	0	0.00	111	22	0.11
184	5	0.16	173	0	0.00
145	11	0.15	177	0	0.09
172	0	0.00	180	19	0.10
133	8	0.11	172	0	0.00
157	21	0.14	181	0	0.10
175	0	0.00	138	14	0.10

3.16 Normal equations. The theory of the previous chapter applies; a multiple regression function of the linear form

$$Y_r = a + bU + cV$$

has two degrees of freedom, i.e., as many degrees of freedom as there are independent variables. The residual variance has $n - 3$ degrees of freedom, three degrees of freedom being lost by the calculation of a (or $\bar{Y}$), b, or c from the data. If this regression plane is to be fitted to the

ungrouped data by the method of least squares, we shall have the following normal equations:

$$an + b\sum U + c\sum V \quad = \sum Y$$
$$a\sum U + b\sum U^2 + c\sum VU = \sum YU$$
$$a\sum V + b\sum VU + c\sum V^2 = \sum YV$$

from which a, b, and c may be found.

We have

$$\begin{aligned}
\sum Y &= 8{,}907 \\
\sum U &= 312 \\
\sum V &= 4.54 \\
\sum VU &= 39.05 \\
\sum YU &= 52{,}335 \\
\sum YV &= 778.41 \\
\sum U^2 &= 5{,}372 \\
\sum V^2 &= 0.5926 \\
n &= 52
\end{aligned}$$

from which

$$\begin{aligned}
a &= 171.258 \\
b &= -0.4133 \\
c &= 28.750
\end{aligned}$$

The analysis of variance follows:

Source of variation	Sum of squares	Degrees of freedom	Mean square
Regression	479.323	2	239.66
Residual	13,935.350	49	284.39
Total	14,414.673	51	...

The sums of squares may be computed from any two of the following:

$$\sum(Y - \bar{Y})^2 = \sum Y^2 - \frac{(\sum Y)^2}{n}$$

$$\sum(Y_r - \bar{Y})^2 = \sum Y_r^2 - n\bar{Y}^2 = a\sum Y + b\sum YU + c\sum YV - \frac{(\sum Y)^2}{n}$$

$$\sum(Y - Y_r)^2 = \sum Y^2 - a\sum Y - b\sum YU - c\sum YV$$

The mean square due to regression is seen to be even less than that not associated with regression. No F test is necessary; clearly the relationship

$$Y = 171.258 - 0.4133U + 28.750V$$

is inadequate.

3.17 Further examples. The following data on the tensile strength, hardness, and density of 60 specimens of die-cast aluminum are given by Shewhart (36).

Tensile strength (pounds per square inch)	Hardness (Rockwell E)	Density (grams per cubic centimeter)	Tensile strength (pounds per square inch)	Hardness (Rockwell E)	Density (grams per cubic centimeter)
29,314	53.0	2.666	29,250	71.3	2.648
34,860	70.2	2.708	27,992	52.7	2.400
36,818	84.3	2.865	31,852	76.5	2.692
30,120	55.3	2.627	27,646	63.7	2.669
34,020	78.5	2.581	31,698	69.2	2.628
30,824	63.5	2.633	30,844	69.2	2.696
35,396	71.4	2.671	31,988	61.4	2.648
31,260	53.4	2.650	36,640	83.7	2.775
32,184	82.5	2.717	41,578	94.7	2.874
33,424	67.3	2.614	30,496	70.2	2.700
37,694	69.5	2.524	29,668	80.4	2.583
34,876	73.0	2.741	32,622	76.7	2.668
24,660	55.7	2.619	32,822	82.9	2.679
34,760	85.8	2.755	30,380	55.0	2.609
38,020	95.4	2.846	38,580	83.2	2.721
25,680	51.1	2.575	28,202	62.6	2.678
25,810	74.4	2.561	29,190	78.0	2.610
26,460	54.1	2.593	35,636	84.6	2.728
28,070	77.8	2.639	34,332	64.0	2.709
24,640	52.4	2.611	34,750	75.3	2.880
25,770	69.1	2.696	40,578	84.8	2.949
23,690	53.5	2.606	28,900	49.4	2.669
28,650	64.3	2.616	34,648	74.2	2.624
32,380	82.7	2.748	31,244	59.8	2.705
28,210	55.7	2.518	33,802	75.2	2.736
34,002	70.5	2.726	34,850	57.7	2.701
34,470	87.5	2.875	36,690	79.3	2.776
29,248	50.7	2.585	32,344	67.6	2.754
28,710	72.3	2.547	34,440	77.0	2.660
29,830	59.5	2 606	34,650	74.8	2.819

Is the multiple regression of tensile strength (Y) on hardness (U) and density (V) significant?

Source of variation	Sum of squares	Degrees of freedom	Mean square
Regression	524,681,187	2	262,340,593.5
Residual	417,700,935	57	7,328,086
Total	942,274,585	59	...

$F = \dfrac{262{,}340{,}594}{7{,}328{,}086} = 35.80$. From Table VIII, for 2 and 57 degrees of freedom, we need $F = 4.98$ for highly significant regression. Hence the equation found by Shewhart

$$Y_r = 150.988U + 15310.35V$$

describes the relationship between tensile strength, hardness, and density.

In the preceding chapter, it was shown that a linear relationship between the life of light bulbs and a certain test of filament wire was not statistically significant. A second type of test was made on the wire. Jennett and Dudding (24) report the following results, the first two columns showing the data already considered on page 103.

Life of bulbs Y	Test of wire U	Test of wire V
1605	276	14.2
1120	293	15.6
1320	288	16.1
1225	315	—
1055	305	15.2
1390	315	14.6
1385	306	21.4
1700	286	19.4
2070	289	18.9
1395	296	18.5
1105	335	20.8

Would a multiple relation of the form

$$Y_r = a + bU + cV$$

succeed where the linear relation between Y and U failed?

We find

$$a = 4{,}336.29$$
$$b = -12.69$$
$$c = 49.80$$

Source of variation	Sum of squares	Degrees of freedom	Mean square
Regression	388,795.13	2	194,397.57
Residual	481,227.37	7	68,746.77
Total	870,022.50	9	...

$F = \dfrac{194{,}397.57}{68{,}746.77} = 2.83$. From Table VIII, for 2 and 7 degrees of freedom, a value of $F = 4.74$ is required for significance at the 0.05 level. The regression is not statistically significant even when two independent variables are included.

NOTES

3.18 Least squares and multiple regression. In fitting a regression plane

$$Y_r = a + bU + cV$$

to observed data by the method of least squares,

$$\sum(Y - Y_r)^2$$

is required to be minimum, Y being the observed values of the dependent variable. Write

$$\varphi = \sum(Y - a - bU - cV)^2$$

For φ minimum

$$\frac{\partial\varphi}{\partial a} = -2\sum(Y - a - bU - cV) = 0$$

[6] $$\frac{\partial\varphi}{\partial b} = -2\sum(Y - a - bU - cV)U = 0$$

$$\frac{\partial\varphi}{\partial c} = -2\sum(Y - a - bU - cV)V = 0$$

Furthermore $\partial^2\varphi/\partial a^2$, $\partial^2\varphi/\partial b^2$, $\partial^2\varphi/\partial c^2$ are positive. Rearranging [6] we have the normal equations

$$\sum Y = na + b\sum U + c\sum V$$

$$\sum YU = a\sum U + b\sum U^2 + c\sum VU$$

$$\sum YV = a\sum V + b\sum VU + c\sum V^2$$

which may be solved simultaneously for a, b, and c.

3.19 Computation of the sums of squares for multiple regression analysis. In order to reduce the labor involved in computing sums of squares, the following were suggested:

$$\sum(Y - \bar{Y})^2 = \sum Y^2 - n\bar{Y}^2$$

$$\begin{aligned}\sum(Y_r - \bar{Y})^2 &= \sum(Y_r - \bar{Y})(Y_r - \bar{Y}) = \sum(Y_r - \bar{Y})(Y_r)\\ &= \sum(a + bU + cV - \bar{Y})(Y_r)\\ &= a\sum Y_r + b\sum UY_r + c\sum VY_r - \bar{Y}\sum Y_r\\ &= a\sum Y + b\sum UY + c\sum VY - \left(\frac{\sum Y}{n}\right)^2\end{aligned}$$

since

$$\sum Y_r = \sum(a + bU + cV) = na + b\sum U + c\sum V = \sum Y$$

and similarly for $\sum UY_r$ and $\sum VY_r$.

$$\begin{aligned}\sum(Y - Y_r)^2 &= \sum(Y - Y_r)(Y - Y_r)\\ &= \sum(Y - Y_r)Y - \sum(Y - Y_r)Y_r\\ &= \sum(Y - Y_r)Y - \sum(Y - Y_r)(a + bU + cV)\end{aligned}$$

From the three normal equations, we have

$$\sum(Y - Y_r) = 0, \quad \sum(Y - Y_r)U = 0, \quad \sum(Y - Y_r)V = 0$$

Hence

$$\begin{aligned}\sum(Y - Y_r)^2 &= \sum(Y - Y_r)Y\\ &= \sum(Y - a - bU - cV)Y\\ &= \sum Y^2 - a\sum Y - b\sum UY - c\sum VY\end{aligned}$$

3.20 Analysis of covariance. Furry (19) gives the following data on the breaking strength and thickness of starch films.

BREAKING STRENGTH AND THICKNESS OF STARCH FILMS*

Film	Thickness (inches $\times 10^{-4}$)	Breaking strength (grams)
Wheat starch		
1	5.0	263.7
2	3.5	130.8
3	4.7	382.9
4	4.3	302.5
5	3.8	213.3
6	3.0	132.1
7	4.2	292.0
8	4.5	315.5
9	4.3	262.4
10	4.1	314.4
11	5.5	310.8
12	4.8	280.0
13	4.8	331.7
14	8.0	672.5
15	7.4	496.0
16	5.2	311 9
17	4.7	276.7
18	5.4	325.7
19	5.4	310.8
20	5.4	288.0
21	4.9	269.3
Dasheen starch		
1	7.0	485.4
2	6.0	395.4
3	7.1	465.4
4	5.3	371.4
5	6.2	402.0
6	5.8	371.9
7	6.6	430.0
8	6.6	380.0

Film	Thickness (inches $\times 10^{-4}$)	Breaking strength (grams)
Rice starch		
1	7.1	556.7
2	6.7	552.5
3	5.6	397.5
4	8.1	532.3
5	8.7	587.8
6	8.3	520.9
7	8.4	574.3
8	7.3	505.0
9	8.5	604.6
10	7.8	522.5
11	8.0	555.0
12	8.4	561.1
Sweet potato starch		
1	9.4	837.1
2	10.6	901.2
3	9.0	595.7
4	7.6	510.0
Canna starch		
1	7.7	791.7
2	6.3	610.0
3	8.6	710.0
4	11.8	940.7
5	12.4	990.0
6	12.0	916.2
7	11.4	835.0
8	10.4	724.3
9	9.2	611.1
10	9.0	621.7
11	9.5	735.4
12	12.5	990.0
13	11.7	862.7

* Each value for thickness and breaking strength is the average of five or more film-strip tests.

Corn starch			Potato starch		
1	8.0	731.0	1	13.0	983.3
2	7.3	710.0	2	13.3	958.8
3	7.2	604.7	3	10.7	747.8
4	6.1	508.8	4	12.2	866.0
5	6.4	393.0	5	11.6	810.8
6	6.4	416.0	6	9.7	950.0
7	6.9	400.0	7	10.8	1,282.0
8	5.8	335.6	8	10.1	1,233.8
9	5.3	306.4	9	12.7	1,660.0
10	6.7	426.0	10	9.8	746.0
11	5.8	382.5	11	10.0	650.0
12	5.7	340.8	12	13.8	992.5
13	6.1	436.7	13	13.3	896.7
14	6.2	333.3	14	12.4	873.9
15	6.3	382.3	15	12.2	924.4
16	6.0	397.7	16	14.1	1,050.0
17	6.8	619.1	17	13.7	973.3
18	7.9	857.3	....	....	.....
19	7.2	592.5	....	....	.....

Analysis of variance of breaking strengths

Source of variation	Sum of squares	Degrees of freedom	Mean square
Among starches	5,307,433.08	6	884,572.18
Within starches	1,987,918.13	87	22,849.63
Total	7,295,351.21	93	...

The differences in breaking strengths from starch to starch are highly significant. But examination of the data indicates that at least some of this apparent significance is due merely to differences in the thickness of the starch film and not to any chemical superiority of certain starches over others. To determine the effect of thickness on strength, the relationships between the two must first be measured for each starch; for our purposes the best measure is given by the regression coefficient b (the slope of the regression line) of breaking strength on thickness.

$$b = \frac{\sum(Y - \bar{Y})(X - \bar{X})}{\sum(X - \bar{X})^2} = \frac{\sum yx}{\sum x^2} = \frac{\sum YX - n\bar{Y}\bar{X}}{\sum X^2 - n\bar{X}^2}$$

where Y represents a breaking strength value and X represents the corresponding value of thickness. To illustrate, we have for sweet-potato starch

$$b = \frac{26{,}658.76 - 26{,}022.60}{339.48 - 334.89}$$

$$= 138.60$$

We form the following table:

Starch	$\sum y^2$	$\sum yx$	$\sum x^2$	b	$\sum y^2 - \frac{(\sum xy)^2}{\sum x^2}$
Wheat	254,104.85	2,310.53	25.96	89.00	48,459.67
Dasheen	13,215.27	156.39	2.65	59.02	3,985.90
Corn	447,055.68	1,866.60	9.88	188.93	94,404.31
Rice	30,993.44	417.61	9.15	45.64	11,933.54
Sweet potato	105,772.34	636.16	4.59	138.60	17,602.50
Canna	232,013.75	2,763.55	46.61	59.29	68,160.32
Potato	904,762.80	1,192.81	36.66	32.54	865,952.22
Total	1,987,918.13	9,343.65	135.50		1,110,498.46

We should like to eliminate from the variation in breaking strength that part attributable to variation in thickness; to do so it would be convenient to use an average within-starch regression coefficient such as would be given by $b = 9{,}343.65/135.50$. This is permissible only if the differences among the seven regression coefficients are not statistically significant. To test the latter, we have

Source of variation	Sum of squares	Degrees of freedom	Mean square
Deviation of within-starch regression lines from average within-starch regression line $\sum(Y_r - Y_{\bar{r}})^2$	233,111.22	6	38,851.87
Deviation of observations from within-starch regression (error) $\sum(Y - Y_r)^2$	1,110,498.46	80	13,881.23
Deviation of observations from average within-starch regression line $\sum(Y - Y_{\bar{r}})^2$	1,343,609.68	86	

The calculation of the above sums of squares can be carried out in the following way:

$$\sum(Y - Y_{\bar{r}})^2 = \sum y^2 - \frac{(\sum yx)^2}{\sum x^2}$$

The value of $\sum y^2$ for within starches for the entire table is 1,987,918.13. Also

$$\frac{(\sum yx)^2}{\sum x^2} = \frac{(9{,}343.65)^2}{135.50} = 644{,}308.45$$

or

$$\sum(Y - Y_{\bar{r}})^2 = 1{,}343{,}609.68$$

In $\sum(Y - Y_r)^2$, deviations are measured from the within-starch regression lines, i.e.,

$$\sum(Y - Y_r)^2 = 1{,}110{,}498.46$$

By subtraction

$$\sum(Y_r - Y_{\bar{r}})^2 = 233{,}111.22$$

The allocation of degrees of freedom is easily explained. For the total sum of squares, $\sum(Y - Y_{\bar{r}})^2$, there are the 87 degrees of freedom allocated to within-starch variation in the original analysis of variance less the 1 degree of freedom lost by the fact that the deviations are taken about the over-all within-regression line. Second, from the original 87 degrees for within-starch variation we must subtract 7 degrees of freedom attributable to the 7 regression lines, leaving 80 degrees of freedom for the term $\sum(Y - Y_r)^2$. By subtraction, there remain 6 degrees of freedom for $\sum(Y_r - Y_{\bar{r}})^2$.

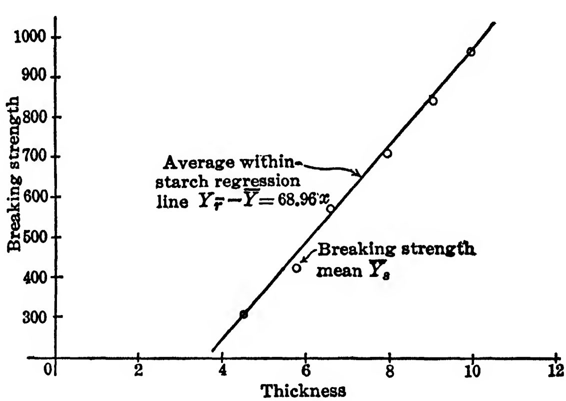
Applying the F test, we find

$$F = \frac{38{,}851.87}{13{,}881.23} = 2.80$$

which for 6 and 80 degrees of freedom is significant at the 5 per cent level but not significant at the 1 per cent level. It is up to the analyst to decide whether or not he will continue. We shall consider the regression coefficients to be not significantly different; they are presumed to be replaced by

$$b = \frac{9{,}343.65}{135.50} = 68.96$$

We are now able to calculate adjusted breaking strength means for each starch, each strength mean being corrected by elimination of the effect of thickness. The figure below illustrates the situation. Clearly most of the



variation in breaking strength is explained by variation in thickness. We have, for any mean $\bar{Y}_s$

$$\bar{Y}_s = \bar{Y} + \underbrace{(\bar{Y}_s - Y_{\bar{r}})}_{\text{Variation around regression}} + \underbrace{(Y_{\bar{r}} - \bar{Y})}_{\text{Variation due to regression}}$$

We wish to eliminate the effect of the last term on $\bar{Y}_s$. Hence we want corrected values of $\bar{Y}_s$, given by

$$\text{Corrected } \bar{Y}_s = \bar{Y} + (\bar{Y}_s - \bar{Y}_{\bar{r}})$$

or since

$$Y_{\bar{r}} - \bar{Y} = bx$$

$$\text{Corrected } \bar{Y}_s = \bar{Y}_s - bx$$

The following tabular form is convenient.

Starch	Original mean breaking strength $\bar{Y}_s$	Mean thickness $\bar{X}_s$	$\bar{X}_s - \bar{X}$	$bx = b(\bar{X}_s - \bar{X})$	Corrected mean breaking strength
Wheat	308.7	4.90	−3.00	−206.9	515.6
Dasheen	412.7	6.33	−1.57	−108.3	521.0
Corn	482.8	6.53	−1.37	− 94.5	577.3
Rice	539.2	7.74	−0.16	− 11.0	550.2
Sweet potato	711.0	9.15	1.25	86.2	624.8
Canna	795.3	10.19	2.29	157.9	637.4
Potato	976.5	11.96	4.06	280.0	696.5

We can now judge the effect of the independent variable, thickness. The original analysis of variance is replaced by the following table:

Source of variation	Sum of squares	Degrees of freedom	Mean square
Among starches	99,947.77	6	16,657.96
Within starches	1,343,609.68	86	15,623.37
Total	1,443,557.45	92	

The total sum of squares 1,443,557.45 differs from the previous total 7,295,351.21, for the latter represented variation about the grand mean $\bar{Y}$, i.e.,

$$\sum(Y - \bar{Y})^2$$

whereas the present total sum of squares represents variation of the observations about a regression line fitting the 94 points, i.e.,

$$\sum(Y - Y_R)^2$$

The same is true for the present within-starch sum of squares. The among-starch sum of squares is obtained by subtraction. The necessity of computing these terms makes it desirable to set up at the outset a table of the form shown herewith.

Source of variation	$\sum y^2$	$\sum yx$	$\sum x^2$
Among starches	5,307,433.09	56,268.32	600.16
Within starches	1,987,918.13	9,343.65	135.50
Total	7,295,351.22	65,611.97	735.66

Using the equation given on page 122, we have

$$7{,}295{,}351.22 - \frac{(65{,}611.97)^2}{735.66} = 1{,}443{,}557.45$$

$$1{,}987{,}918.13 - \frac{(9{,}343.65)^2}{135.50} = 1{,}343{,}609.68$$

and the value 99,947.77 is obtained by subtraction. The degrees of freedom differ from the original table only in that there are 86 instead of 87 degrees of freedom for the within-starch (error) term, for now error is measured by variation about regression rather than variation about starch means, and 1 additional degree of freedom is lost by the calculation of the regression coefficient from the data.

An F test yields

$$F = \frac{16{,}657.96}{15{,}623.37} = 1.066$$

which, for 6 and 86 degrees of freedom, is not significant. The differences in breaking strengths are attributable to differences in thicknesses and not to kinds of starches. Judgment on the basis of the original analysis of variance might have been misleading; the inclusion of the independent variable, thickness, was essential if proper conclusions were to be reached.

CHAPTER IV

SYSTEMATIC QUALITY CONTROL

4.1 Introduction. The preceding chapters considered the design and analysis of industrial experiments which aim to identify the factors responsible for variable quality. The present chapter describes the contribution to this objective of a routinized system of recording quality data.

In this chapter our objective is, essentially, to judge the quality of current output against standards. The standards may be set by technical commissions, by government, or they may represent the quality of the product during the past. If current output, as known from current samples of information, departs from the standard by an amount which is statistically significant, an economic loss may be involved. Output, the quality of which is significantly higher than intended, implies wastage of labor and materials; the dis-economies of lower than standard quality are equally obvious. If possible, the responsible factor or factors should be immediately identified and removed.

Standards formed from past experience may be based on the quality records of a fixed period of time during the past or on a period which changes as time goes on in order to incorporate the records of the more recent past. There are also variations of these schemes. We shall illustrate only the case in which the time period is increasing. That is, after the quality of the output of say the 50th day is judged against the standard of the previous 49 days, the data of the 50th day is added to the previous population, and the quality of the 51st day is judged by comparison with the standards based on 50 days of data. The method of handling fixed or other varieties of shifting standards involves only minor changes in the following discussion and the reader can supply them for himself.

Standards based on accumulated data are valid only if those data are homogeneous. Thus a standard in the form of a mean of say 20 weeks' data has little sense if the 20 weekly means differ significantly among themselves. We shall test each population for such homogeneity. In order to preserve the homogeneity of shifting populations it has occasionally been the practice not to add to the population without adjust-

ment any data on current output which departs significantly from the standard. This practice has no statistical justification and is not recommended.

Finally the populations must be large relative to the size of the current samples. This is clearly sensible from a practical point of view; moreover it is important statistically if we are to assume exact knowledge of such population parameters as $\bar{X}'$ and σ, which are or form the basis of industrial standards.

4.2 Population: formation and homogeneity. Supplement B of the American Society for Testing Materials' " Manual on Presentation of Data " (1) gives the following information on an operating characteristic:

Sample number	Sample size n	Mean quality $\bar{X}$	Standard deviation s
1	50	35.7	5.35
2	50	34.6	5.03
3	50	32.6	3.43
4	50	35.3	4.55
5	50	33.4	4.10
6	50	35.2	4.30
7	50	33.3	5.18
8	50	33.9	5.30
9	50	32.3	3.09
10	50	33.7	3.67

From this information we want to estimate the mean $\bar{X}'$ and the standard deviation σ of the population of 500 observations formed by combining these 10 samples.

For k samples of size $n_1, \cdots, n_k$ with respective means $\bar{X}_1, \cdots, \bar{X}_k$ and respective variances $s_1^2, \cdots, s_k^2$, we know that

$$\bar{X}' = \frac{\sum n_i \bar{X}_i}{\sum n_i}$$

and

$$\hat{\sigma}^2 = \frac{\sum n_i s_i^2}{\sum n_i - k}$$

In the present example $\bar{X}' = 34.0$

and $\hat{\sigma} = 4.51$

Is this population homogeneous in its mean? Assuming normality, approximately 90 per cent of the sample means (i.e., nine means) should

Sample number	Sample size n	Total of observed data on Btu content	Sample mean $\bar{X}$	Total observations to date $\sum n$	Total of observed data on Btu content to date	Mean to date $\bar{X}'$	$\sum (X-\bar{X})^2$ for each sample	s Sample standard deviation $= \sqrt{\frac{\Sigma(X-\bar{X})^2}{n}}$	$\sum (X-\bar{X})^2$ for all data to date	Number of observations to date, less number of samples $\sum n - k$	Estimated standard deviation to date $\hat{\sigma}$	Range
1	14	7,523	537.36	14	7,523	537.36	300.930	4.64	300.930	13	4.81	17
2	14	7,604	543.14	28	15,127	540.25	231.160	4.06	532.090	26	4.52	15
3	14	7,530	537.86	42	22,657	539.45	201.720	3.80	733.810	39	4.34	15
4	14	7,496	535.43	56	30,153	538.45	173.430	3.52	907.240	52	4.18	11
5	14	7,532	538.00	70	37,685	538.36	86.000	2.48	993.240	65	3.91	9
6	14	7,518	537.00	84	45,203	538.13	204.000	3.82	1,197.240	78	3.92	13
7	14	7,489	534.93	98	52,692	537.67	64.930	2.15	1,262.170	91	3.72	8
8	14	7,486	534.71	112	60,178	537.30	108.856	2.79	1,371.026	104	3.63	9
9	14	7,541	538.64	126	67,719	537.45	177.220	3.56	1,548.246	117	3.64	15
10	14	7,631	545.07	140	75,350	538.21	216.930	3.94	1,765.176	130	3.68	11
11	14	7,594	542.43	154	82,944	538.60	269.430	4.39	2,034.606	143	3.77	18
12	14	7,570	540.71	168	90,514	538.77	16.856	1.10	2,051.462	156	3.63	3
13	14	7,565	540.36	182	98,079	538.90	147.220	3.24	2,198.682	169	3.61	12
14	14	7,541	538.64	196	105,620	538.88	196.220	3.74	2,394.902	182	3.63	11
15	14	7,573	540.93	210	113,193	539.01	84.930	2.46	2,479.832	195	3.57	10
16	14	7,587	541.93	224	120,780	539.20	160.930	3.39	2,640.762	208	3.56	12
17	14	7,563	540.21	238	128,343	539.26	248.356	4.21	2,889.118	221	3.62	13
18	14	7,534	538.14	252	135,877	539.19	107.720	2.77	2,996.838	234	3.58	10
19	14	7,532	538.00	266	143,409	539.13	223.000	3.99	3,219.838	247	3.61	13
20	14	7,529	537.79	280	150,938	539.06	138.356	3.14	3,358.194	260	3.59	13
21	14	7,525	537.50	294	158,463	538.99	185.500	3.64	3,543.694	273	3.60	13
22	14	7,509	536.36	308	165,972	538.87	77.220	2.35	3,620.914	286	3.56	9

23	14	7,533	538.07	322	173,505	538.84	124.930	2.99	3,745.844	299	3.54	12
24	14	7,535	538.21	336	181,040	538.81	156.356	3.34	3,902.200	312	3.54	12
25	14	7,517	536.93	350	188,557	538.73	128.830	3.03	4,031.030	325	3.52	11
26	14	7,532	538.00	364	196,089	538.71	70.000	2.24	4,101.030	338	3.48	8
27	14	7,506	536.14	378	203,595	538.61	131.720	3.07	4,232.750	351	3.47	11
28	14	7,494	535.29	392	211,089	538.49	44.856	1.79	4,277.606	364	3.43	5
29	14	7,482	534.43	406	218,571	538.35	67.430	2.19	4,345.036	377	3.39	7
30	14	7,505	536.07	420	226,076	538.28	68.930	2.22	4,413.966	390	3.36	6
31	14	7,538	537.71	434	233,604	538.26	112.856	2.84	4,526.822	403	3.35	10
32	14	7,504	536.00	448	241,108	538.19	110.000	2.80	4,636.822	416	3.34	9
33	14	7,529	537.79	462	248,637	538.18	104.356	2.73	4,741.178	429	3.32	11
34	14	7,482	534.43	476	256,119	538.07	55.430	1.99	4,796.608	442	3.29	6
35	14	7,517	536.93	490	263,636	538.03	100.930	2.69	4,897.538	455	3.28	9
36	14	7,520	537.14	504	271,156	538.01	25.720	1.36	4,923.258	468	3.24	5
37	14	7,522	537.29	518	278,678	537.99	156.856	3.35	5,080.114	481	3.25	12
38	14	7,506	536.14	532	286,184	537.94	125.720	3.00	5,205.834	494	3.25	12
39	14	7,521	537.21	546	293,705	537.92	148.356	3.26	5,354.190	507	3.25	11
40	14	7,526	537.57	560	301,231	537.91	105.430	2.74	5,459.620	520	3.24	10
41	14	7,544	538.86	574	308,775	537.94	80.920	2.40	5,540.540	533	3.22	8
42	14	7,480	534.29	588	316,255	537.85	160.856	3.39	5,701.396	546	3.23	12
43	14	7,510	536.43	602	323,765	537.82	111.430	2.82	5,812.826	559	3.22	10
44	14	7,526	537.57	616	331,291	537.81	127.430	3.02	5,940.256	572	3.22	10
45	14	7,507	536.21	630	338,798	537.77	50.356	1.90	5,990.612	585	3.20	7
46	14	7,531	537.93	644	346,329	537.78	198.930	3.77	6,189.542	598	3.22	14
47	14	7,489	534.93	658	353,818	537.72	158.930	3.37	6,348.472	611	3.22	11
48	14	7,482	534.43	672	361,330	537.65	217.430	3.94	6,565.902	624	3.24	12
49	14	7,511	536.50	686	368,811	537.63	179.500	3.58	6,745.402	637	3.25	12
50	14	7,504	536.00	700	376,315	537.59	148.000	3.25	6,893.402	650	3.26	11

fall within $\bar{X}' \pm 1.65\sigma_{\bar{X}}$, where $\sigma_{\bar{X}} = \hat{\sigma}/\sqrt{n}$. We have

$$\sigma_{\bar{X}} = \frac{4.51}{\sqrt{50}} = 0.638$$

Only five means fall within $\bar{X}' \pm 1.65\sigma_{\bar{X}}$. The population cannot be said to be homogeneous in its mean.

Possibly the lack of homogeneity of the population is due to significant differences among the 10 sample standard deviations. The distribution of s for random samples from a normal population was given on page 46. For large samples, say $n > 30$, it is easy to show that this distribution is approximately normal with standard error σ_s equal to $\sigma/\sqrt{2n}$. We have

$$\sigma_s = \frac{4.51}{\sqrt{100}} = 0.451$$

Five values of s fall outside the range $\hat{\sigma} \pm 1.65\sigma_s$; the sample standard deviations differ significantly among themselves. The parameters of this population ($\bar{X}'$ and σ) could not be effectively used as standards with which to compare current quality.

4.3 Example involving $\bar{X}$. The data in the first three columns of the preceding table have been reported by Pettebone and Young (32). They cover 50 *consecutive* samples each of 14 observations; the quality characteristic is the Btu value of a mixed fuel gas.

The functions of the various columns will be discussed presently. The tabular form used was originally given by Dudding and Baker (10).

In order to permit the reader to check all entries in the first three rows of the preceding table, the individual observations of the first three samples are now given.

Btu value

Sample 1	Sample 2	Sample 3
533	546	541
537	540	534
535	542	528
540	550	538
542	541	542
547	547	540
543	546	539
536	539	538
531	542	543
530	535	533
536	542	538
534	542	537
538	544	540
541	549	539

4.4 The range. In the last column the range, which is the difference between the largest and smallest observations, is recorded. The principal merit of the range lies in the ease with which it is computed, and its utility arises from the fact that for a normal population the mean range over k samples (k large) bears a fixed relationship to the more common measure of variation, the standard deviation σ of the population. For $k = 50$, the mean range will be found to be 10.68. For sample size $n = 14$, we find from Table VI

$$\frac{10.68}{\hat{\sigma}} = 3.40676$$

or

$$\hat{\sigma} = 3.135$$

which is in good agreement with $\hat{\sigma} = 3.26$ found by the more efficient method that is illustrated on page 127. For small k, say less than 10, or large n, say more than 15, the mean range method of estimating σ is unreliable. In our example, the agreement between the two methods of estimating σ happens to be good even for small k. Thus for $k = 2$, the two methods of estimating yield 4.52 and 4.70; for $k = 5$, they yield 3.19 and 3.93.

The columns of the large table contain the information needed for a simple quality control record. Thus, at the end of the 20th week, the mean and standard deviation of the accumulated population are

$$\bar{X}' = 539.06$$

and

$$\hat{\sigma} = 3.59$$

We wish to determine whether or not the mean of the sample of the 21st week (based on n observations) falls within $\pm 2\sigma_{\bar{X}}$ of $\bar{X}'$; where

$$\sigma_{\bar{X}} = \frac{\hat{\sigma}}{\sqrt{n}} = \frac{3.59}{\sqrt{14}} = 0.96$$

It does not; hence the quality of the output of the 21st week does not conform to the standard ($\bar{X}' = 539.06$). If records of influential factors, such as kind of coal burned, are kept simultaneously, the cause of lack of control can often be immediately spotted and corrected.

The steps of this procedure is systematized in the following table. The calculations begin with the 20th sample so that the beginning population will be 20 times as large as the first sample to be judged.

Sample number	$\bar{X}'$	$\hat{\sigma}$	n size of following sample	$\sigma_{\bar{X}} = \frac{\hat{\sigma}}{\sqrt{n}}$	$\bar{X}' - 2\sigma_{\bar{X}}$	$\bar{X}' + 2\sigma_{\bar{X}}$	$\bar{X}$ of following sample	Under control
20	539.06	3.59	14	0.96	537.14	540.98	537.50	Yes.
21	538.99	3.60	14	0.96	537.07	540.91	536.36	No.
22	538.87	3.56	14	0.95	536.97	540.77	538.07	Yes.
23	538.84	3.54	14	0.95	536.94	540.74	538.21	Yes.
24	538.81	3.54	14	0.95	536.91	540.71	536.93	Yes.
25	538.73	3.52	14	0.94	536.85	540.61	538.00	Yes.
26	538.71	3.48	14	0.93	536.85	540.57	536.14	No.
27	538.61	3.47	14	0.93	536.75	540.47	535.29	No.
28	538.49	3.43	14	0.92	536.65	540.33	534.43	No.
29	538.35	3.39	14	0.91	536.53	540.17	536.07	No.
30	538.28	3.36	14	0.90	536.48	540.08	537.71	Yes.
31	538.26	3.35	14	0.90	536.46	540.06	536.00	No.
32	538.19	3.34	14	0.89	536.41	539.97	537.79	Yes.
33	538.18	3.32	14	0.89	536.40	539.96	534.43	No.
34	538.07	3.29	14	0.88	536.31	539.83	536.93	Yes.
35	538.03	3.28	14	0.88	536.27	539.79	537.14	Yes.
36	538.01	3.24	14	0.87	536.27	539.75	537.29	Yes.
37	537.99	3.25	14	0.87	536.25	539.73	536.14	No.
38	537.94	3.25	14	0.87	536.20	539.68	537.21	Yes.
39	537.92	3.25	14	0.87	536.18	539.66	537.57	Yes.
40	537.91	3.24	14	0.87	536.17	539.65	538.86	Yes.
41	537.94	3.22	14	0.86	536.22	539.66	534.29	No.
42	537.85	3.23	14	0.86	536.13	539.57	536.43	Yes.
43	537.82	3.22	14	0.86	536.10	539.54	537.57	Yes.
44	537.81	3.22	14	0.86	536.09	539.53	536.21	Yes.
45	537.77	3.20	14	0.86	536.05	539.49	537.93	Yes.
46	537.78	3.22	14	0.86	536.06	539.50	534.93	No.
47	537.72	3.22	14	0.86	536.00	539.44	534.43	No.
48	537.65	3.24	14	0.87	535.91	539.39	536.50	Yes.
49	537.63	3.25	14	0.87	535.89	539.37	536.00	Yes.
50	537.59	3.26	..	...	......	......	......	...

The judgments in the last column are valid only if the population against which the sample is being tested is homogeneous. At each stage of the process the homogeneity of the population should be tested. Thus at the end of 20 weeks

$$\sigma_{\bar{X}} = \frac{\hat{\sigma}}{\sqrt{n}} = \frac{3.59}{\sqrt{14}} = 0.96$$

$$\bar{X}' = 539.06$$

$$\bar{X}' - 2\sigma_{\bar{X}} = 537.14$$

$$\bar{X}' + 2\sigma_{\bar{X}} = 540.98$$

In 8 out of 20 samples constituting this population (which is far greater than the 5 per cent of 20 — or 1 sample — which could be attributed to chance) the sample mean $\bar{X}$ fell outside the limits

$$\bar{X}' \pm 2\sigma_{\bar{X}}$$

This lack of control is possibly attributable to significant differences among the sample standard deviations s_i. Very roughly, for n as small as 14, we have

$$\sigma_s = \frac{\hat{\sigma}}{\sqrt{2n}} = \frac{3.59}{\sqrt{28}} = 0.679$$

Two of 20 values of s_i fall outside $\hat{\sigma} \pm 2\sigma_s$, only 1 above the allowable limit (1 in 20).

We conclude that the population formed of the first 20 samples is clearly not homogeneous and the judgment of the mean quality of the 21st sample given in the preceding table cannot properly be made. More often than not, in industrial practice, population homogeneity will be achieved only after months of effort, and a quality control program of the kind suggested in this chapter will not be immediately possible. In such cases the statistician can best serve by assisting in the design and analysis of experiments which aim to identify the causes of non-homogeneity.

4.5 Example involving fraction defective p. A similar procedure is available wherever quality must be recorded simply as acceptable or not. The first three columns of the following table exhibit data, recorded by Shoumatoff (37), covering defects found in the primary inspection of standard radio tubes.

The most important population parameters are the fraction defective $\bar{p}$ and the standard deviation σ_p, and these will constitute the standards in the quality control program. Later in this chapter it will be shown that

$$\bar{p} = \frac{\sum pn}{\sum n}$$

$$\sigma_p = \sqrt{\frac{\bar{p}\bar{q}}{n}}$$

p is the fraction defective in a sample of size n, i.e., pn is the number of defects in a sample, $\bar{p}$ is the population fraction defective, and $q = 1 - p$, $\bar{q} = 1 - \bar{p}$.

The various columns of the following table show the necessary sample statistics together with constantly revised estimates of the

population parameters. Inasmuch as n differs from sample to sample, we record $\bar{p}\bar{q}$ rather than $\frac{\bar{p}\bar{q}}{n}$.

Date	Number of tubes inspected n	Number of tubes rejected pn	Fraction defective in sample (in %) p	Total number of tubes inspected to date $\sum n$	Total defects to date $\sum pn$	$\bar{p}$ to date (in %) $= \frac{\sum pn}{\sum n}$	$\bar{p}\bar{q}$
1	16,484	2,008	12.2	16,484	2,008	12.2	1071
2	24,708	2,719	11.0	41,192	4,727	11.5	1018
3	27,599	2,691	9.8	68,791	7,418	10.8	963
4	28,545	2,699	9.5	97,336	10,117	10.4	932
5	31,530	3,377	10.7	128,866	13,494	10.5	940
6	8,588	1,100	12.8	137,454	14,594	10.6	948
7	19,574	1,478	7.6	157,028	16,072	10.2	916
8	28,644	2,170	7.8	185,672	18,242	9.8	884
9	29,256	2,214	7.6	214,928	20,456	9.5	860
10	32,605	2,540	7.8	247,533	22,996	9.3	844
11	9,314	750	8.1	256,847	23,746	9.2	835
12	16,163	1,108	6.9	273,010	24,854	9.1	827
13	25,601	1,945	7.6	298,611	26,799	9.0	819
14	22,170	1,690	7.6	320,781	28,489	8.9	811
15	26,462	2,162	8.2	347,243	30,651	8.8	803
16	7,955	671	8.4	355,198	31,322	8.8	803
17	11,908	790	6.6	367,106	32,112	8.7	794
18	23,162	1,641	7.1	390,268	33,753	8.6	786
19	24,154	1,890	7.8	414,422	35,643	8.6	786
20	25,287	1,911	7.6	439,709	37,554	8.5	778
21	4,955	517	10.4	444,664	38,071	8.6	786
22	20,095	1,525	7.6	464,759	39,596	8.5	778

At the end of the 15th day, we have, for the accumulated population to that date

$$\bar{p} = 8.8 \text{ per cent}$$

$$\bar{p}\bar{q} = 803$$

Does the mean quality of the output of the 16th day conform to the standard set from this short period? We have

$$\sigma_p = \sqrt{\frac{\bar{p}\bar{q}}{n}} = \sqrt{\frac{803}{7955}} = 0.32$$

The percentage defective of the 16th week is 8.4, which lies between $\bar{p} \pm 2\sigma_p$. The advance in quality of 0.4 per cent from the standard $\bar{p}$

is reasonably attributable to chance, and no inquiry is warranted. This is not true of the remaining days, all of which show significant departures from the standard.

The final steps of the procedure are shown in the following table.

Date	$\bar{p}$ (per cent)	$\bar{p}\,\bar{q}$	n number of tubes in following sample	$\sigma_p = \sqrt{\frac{\bar{p}\,\bar{q}}{n}}$	$\bar{p} - 2\sigma_p$	$\bar{p} + 2\sigma_p$	p fraction defective in following sample	Under control
15	8.8	803	7,955	0.32	8.2	9.4	8.4	Yes.
16	8.8	803	11,908	0.26	8.3	9.3	6.6	No.
17	8.7	794	23,162	0.19	8.3	9.1	7.1	No.
18	8.6	786	24,154	0.18	8.2	9.0	9.8	No.
19	8.6	786	25,287	0.18	8.2	9.0	7.6	No.
20	8.5	778	4,955	0.40	7.7	9.3	10.4	No.
21	8.6	786	20,095	0.20	8.2	9.0	7.6	No.
22	8.5	778	......	....	...	...	....	...

As in the previous example, the homogeneity of the current population of sample fraction defectives must be examined. This is somewhat laborious for variable n. To test the homogeneity of the population accumulated to the fifteenth day, we have $\bar{p} = 8.8$ per cent, $\bar{q} = 91.2$ per cent and 15 values of σ_p, depending on n. All 15 sample percentage defectives fall outside these limits. There is, therefore, no evidence of control in the production of these tubes during this (far

Date	σ_p (in %)	$\bar{p} - 2\sigma_p$	$\bar{p} + 2\sigma_p$
1	0.22	8.4	9.2
2	0.18	8.4	9.2
3	0.17	8.5	9.1
4	0.17	8.5	9.1
5	0.16	8.5	9.1
6	0.31	8.2	9.4
7	0.20	8.4	9.2
8	0.17	8.5	9.1
9	0.17	8.5	9.1
10	0.16	8.5	9.1
11	0.29	8.2	9.4
12	0.22	8.4	9.2
13	0.18	8.4	9.2
14	0.19	8.4	9.2
15	0.17	8.5	9.1

too brief) 15-day period, and the judgment previously passed on the quality of the tubes of the 16th day must be rescinded. Allocable causes of variability are present and before any effective quality control program can be set up, as many as possible of these causes must be discovered and removed.

A similar test of homogeneity must be carried out on each successive population.

NOTES

4.6 Probability of t defects. Let the probability of a defect be p. Let q be the probability that the piece is good. $p + q = 1$. If a random sample of n pieces is taken, each selection of a piece being independent of all others, what is the probability P_t of obtaining exactly t defective pieces and $n - t$ good pieces?

The first t pieces may be defective and the remainder good. This probability is $p^t q^{n-t}$. But t defective pieces may be obtained in as many ways as one can form combinations of n pieces taken t at a time, namely, C_t^n, where

$$C_t^n = \frac{n!}{t!(n - t)!}$$

Hence the answer is

$$P_t = C_t^n q^{n-t} p^t$$

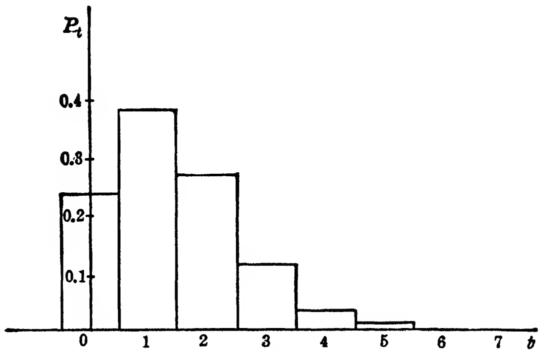
Now $C_t^n q^{n-t} p^t$ is the general expression for the terms of the expansion of $(q + p)^n = 1$, i.e.,

$$1 = (q + p)^n = q^n + nq^{n-1} p + \frac{n(n - 1)}{2!} q^{n-2}p^2 + \cdots$$
$$+ C_t^n q^{n-t}p^t + \cdots + nqp^{n-1} + p^n$$

Hence the successive terms of the above expression give the probability of $0, 1, 2, \cdots, n$ defective pieces. For example, for $p = 1/6$, $q = 5/6$, and $n = 8$, the probabilities are shown in the table below.

t	C_t^n	$q^{n-t}p^t$	$P_t = C_t^n q^{n-t} p^t$
0	1	0.2325680	0.233
1	8	0.0465136	0.372
2	28	0.0093027	0.261
3	56	0.0018605	0.104
4	70	0.0003721	0.026
5	56	0.0000744	0.004
6	28	0.0000149	0.000
7	8	0.0000030	0.000
8	1	0.0000006	0.000
			1.000

The accompanying graph illustrates these results.



4.7 Mean and variance of the fraction defective. We shall calculate the mean and variance of such a distribution of probabilities.

Consider an n-fold experiment ($n = 8$, above). Repeat it N times. Then, for the first moment ${}_0M_1$ around the origin (${}_0M_1$ is the arithmetic mean), we have

${}_0M_1$ = mean number of defects $= \dfrac{f_0 \times 0 + f_1 \times 1 + \cdots + f_n \times n}{N}$ where

f_0 is the frequency of zero defects, f_1 the frequency of 1 defect, etc.

$$
\begin{aligned}
{}_0M_1 &= \frac{Nq^n \times 0 + Nnq^{n-1}p \times 1 + N\dfrac{n(n-1)}{2!}q^{n-2}p^2 \times 2 + \cdots + Np^n n}{N} \\
&= np(q+p)^{n-1} = np
\end{aligned}
$$

To find the variance we first compute the second moment ${}_0M_2$ about zero. The method is due to Bowley (3).

$$
\begin{aligned}
{}_0M_2 &= \frac{f_0 \times 0^2 + f_1 \times 1^2 + \cdots + f_n \times n^2}{N} \\
&= \sum_0^n t^2 C_t^n q^{n-t}p^t \\
&= \sum^n [t(t-1) + t] \times \frac{n(n-1)(n-2)\cdots(n-t+1)}{t!} q^{n-t}p^t \\
&= n(n-1)p^2 \sum \frac{(n-2)!_{t-2}}{(t-2)!} q^{n-t}p^{t-2} + np\sum \frac{(n-1)!_{t-1}}{(t-1)!} q^{n-t}p^{t-1} \\
&= n(n-1)p^2(q+p)^{n-2} + np(q+p)^{n-1} \\
&= n^2p^2 + npq
\end{aligned}
$$

But

$$ {}_0M_2 = \sigma^2 + (\text{mean})^{2*} $$
$$ = \sigma^2 + (np)^2 $$
$$ \therefore \quad \sigma^2 = {}_0M_2 - (np)^2 = n^2p^2 + npq - n^2p^2 = npq $$

In a similar way

$$ \sqrt{\beta_1} = \frac{q-p}{\sqrt{pqn}} $$

and

$$ \beta_2 = 3 + \frac{1-6pq}{pqn} $$

Note that as n increases, $\sqrt{\beta_1} \rightarrow 0$ and $\beta_2 \rightarrow 3$; the distribution of binomial probabilities approaches normality even for $p \neq q$.

To illustrate some of the foregoing results: if the probability of a defective piece in a population is $p = \frac{1}{6}$ and if we draw at random $n = 1000$ pieces, the mean (expected) number of defects is $pn = 167$ and the standard deviation is $\sqrt{pqn} = 11.8$. As the distribution of frequencies of defects is approximately normal for $n = 1000$, we conclude (from Table IV) that in the absence of all " causes " except that of random sampling variation, about 95 per cent of such 1000-observation experiments should have a *frequency* of defects within

$$ pn \pm 2\sqrt{pqn} $$

that is

[1] $$ 167 \pm 23.6 $$

It is, however, generally more useful to record limits on proportions (probabilities) than on frequencies. Each probability value is one nth of the corresponding frequency value; we have

$$ \text{mean fraction defective} = \frac{pn}{n} = p $$

$$ \text{standard deviation of } p = \sigma_p = \frac{\sqrt{pqn}}{n} = \sqrt{\frac{pq}{n}} $$

Thus in about 95 cases in 100 the *proportion* of defects in the 1000-observation experiments should fall between

$$ p \pm 2\sqrt{\frac{pq}{n}} $$

* For any variate φ

$$ \frac{\sum(\varphi - 0)^2}{n} = \frac{\sum(\varphi - \bar{\varphi} + \bar{\varphi} - 0)^2}{n} = \frac{\sum(\varphi - \bar{\varphi})^2}{n} + \frac{\sum(\bar{\varphi} - 0)^2}{n} $$

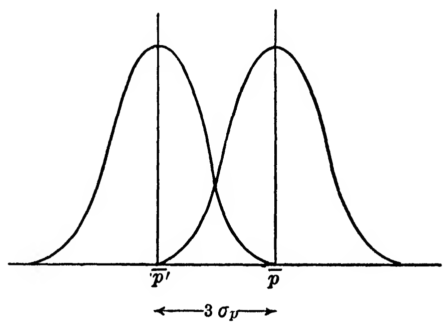
the cross-product being zero. In other symbols ${}_0M_2 = \sigma^2 + \bar{\varphi}^2$.

or

$$\tfrac{1}{6} \pm 0.00236$$

which is, of course, equivalent to [1].

4.8 Limits for control. In setting limits of $\pm 2\sigma_{\bar{X}}$ or $\pm 2\sigma_p$ we expect, even in the ideal case of absence of allocable causes, to find the mean or fraction defective outside these limits in about 5 per cent of the random samples. Thus, as we wish to investigate the reasons for *every* lapse in quality (outside $\pm 2\sigma_{\bar{X}}$ or $\pm 2\sigma_p$) we shall 5 per cent of the time find no allocable cause whatever; for example, the sample fraction defective, though outside $\pm 2\sigma_p$ of $\bar{p}$ actually did occur by chance. If broader limits, say $\pm 3\sigma_p$, are set, the fraction defective will, in the absence of all but chance forces, fall outside these limits in only about 3 of each 1000 random samples. Although we shall then be less frequently searching for " causes " that do not exist, we shall more frequently not be searching for " causes " that may exist. Thus with $\pm 3\sigma_p$ a deviation from $\bar{p}$ so large that it could be expected to happen only once in 100 samples would not be considered to be evidence of lack of control.



This may be stated somewhat differently: if the limits are set at $\pm 3\sigma_p$ and if the true value of the percentage defective $\bar{p}'$ for current output (from which the current sample is drawn) is far off the standard $\bar{p}$, say at $\bar{p} + 3\sigma_p$, then as many as half of all random samples drawn from current output would indicate control, i.e., their percentage defectives would fall within $\pm 3\sigma_p$. With the same out-of-control value for the lot and limits of $\pm 2\sigma_{\bar{p}}$, only 16 per cent of the samples will fall within the control limits.

Limits cannot be set with security until one has accumulated experience as to what limits are economic. It may be suggested that results between $\pm 2\sigma_p$ and $\pm 3\sigma_p$ will bear considerable investigation, whereas results exceeding $\pm 3\sigma_p$ should always receive extensive investigation.

4.9 Notes on σ_p and p. Variation in p from sample to sample, as measured by σ_p, is presumed to be unallocable, i.e., attributable only to the errors of sampling. In industrial data, variation in p from sample to sample is composed not only of such residual errors but of identifiable factors (in our example, differences in workroom humidity). It would not, however, be appropriate to include this element of variability in our estimate of σ_p, for our purpose is to compare the actual variability in p with the variability that would be expected under ideal conditions (random sampling effects only).

p actually varies for another reason, for in industrial practice samples are generally drawn without replacement from a finite population. Assume the population consists of 100,000 tubes, 5 per cent of which are defective. If the

first draw brings forth a bad tube, the probability of a defect is now no longer 0.05 but

$$\frac{4{,}999}{99{,}999}$$

which is not quite 0.05. This point is relevant if we are interested in determining the proportion defective in the batch; but in quality control, where the interest is in spotting absence of control, the population may be considered to be infinite.

CHAPTER V

SAMPLING AND THE RISKS OF PRODUCERS AND BUYERS

5.1 Introduction. A lot of merchandise must often be judged acceptable or not on the basis of information provided by a sample drawn from the lot. In such cases, the producer and buyer will have to incur risks, respectively, of (1) having satisfactory lots rejected and (2) receiving poor lots. If numerical values can be placed on these risks, we may, under certain assumptions to be stated presently, determine the size of sample to be examined and the value of the sample statistic which will differentiate acceptable from non-acceptable lots. Or, if the sample size and the value of the sample statistic are set by authority, the respective risks may be determined.

5.2 Assumptions. In the methods used in this chapter, lots are assumed to be infinite in size relative to the samples drawn from them. It is also assumed that these infinite populations are approximately normal. For example, even though mean quality may decline from $\bar{X}'$ to $\bar{X}''$, the distribution of quality is assumed to remain normal. Finally, the method of sampling from the lots will always be the random method. Further assumptions specific to particular measures of quality will be discussed along with those measures.

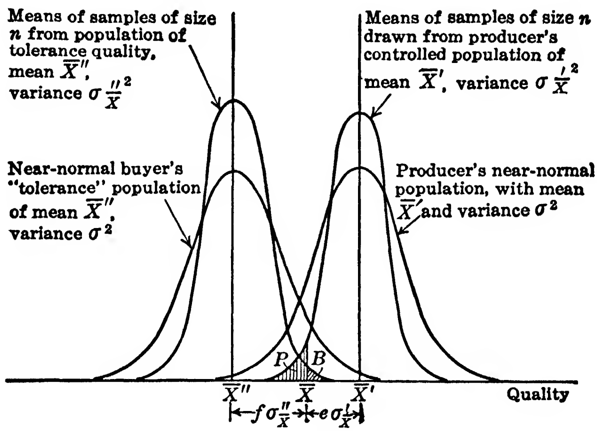
Some experience with the methods of this chapter indicates that these assumptions, while severe, do not prevent effective practical usage of these methods.

This procedure will now be illustrated by four common measures of quality: arithmetic mean, fraction defective, standard deviation, and the coefficient of variation, the latter being the standard deviation divided by the arithmetic mean.

For example, assume that the fraction defective is used as a measure of quality and that the producer's lots average $\bar{p}$. This figure may be acceptable to the buyer who wishes, however, to be protected against the receipt of inferior lots, of quality $\bar{p}_1$ or greater, where $\bar{p}_1 > \bar{p}$, for product defective to this extent will materially affect his operations. The desired specification will state that a sample of n specimens should be drawn from each lot and that lot be marked satisfactory if the sample shows no more than c defective specimens. Such a specification must consider two objectives: first, as already mentioned, the buyer's risk of

receiving highly defective merchandise shall be small (say 1 in 100), and second, the producer's risk of having normally good output (of quality p) rejected shall also be small.

5.3 Producer and buyer risk, using means (Dodge, 9). A lot is to be judged acceptable or not on the basis of the value of the average quality $\bar{X}$ of a sample of n pieces drawn at random from that lot. The producer who is presumed to be manufacturing the product at a statistically controlled or near-controlled plant average quality $\bar{X}'$ would like to run a small risk P of having a lot rejected. The buyer wants to run a small risk B of obtaining lots whose average quality is as low as $\bar{X}''$ or lower. We want to determine $\bar{X}$ and n. The situation is shown graphically below.



The producer's requirements are given by

$$\frac{\bar{X}' - \bar{X}}{\sigma'_{\bar{X}}} = \frac{\bar{X}' - \bar{X}}{\sigma/\sqrt{n}} = e$$

and the buyer's requirements by

$$\frac{\bar{X} - \bar{X}''}{\sigma''_{\bar{X}}} = \frac{\bar{X} - \bar{X}''}{\sigma/\sqrt{n}} = f$$

In addition to the assumptions stated in 5.2, it is assumed here that the lot of "tolerance" quality $\bar{X}''$ has the same variance σ^2 as the lot of usual mean quality $\bar{X}'$. It may also be noted that if the samples are drawn carefully at random, strict normality is not necessary for effective application of the present theorems.

Supplement B of the American Society for Testing Materials' publication "Manual on Presentation of Data" (1) gives the following data

on an operating characteristic. High values indicate high quality. Ranges are recorded rather than the standard deviations, for, as already suggested, ranges are easily calculated and for $n < 15$ a good estimate of σ can be formed from the mean range.

	Number of tests made	Average quality	Range
1	9	37.6	9.5
2	9	31.4	6.0
3	9	34.7	13.5
4	9	35.8	12.0
5	9	38.5	21.0
6	9	34.2	17.5
7	9	36.1	15.5
8	9	32.3	18.0
9	9	35.0	12.5
10	9	33.9	14.0

Assume the producer wants to run no more than 2 chances in 100 that a lot will be rejected and the buyer wants to run no more than 1 chance in 100 that he will receive a lot with average quality less than 30. How many pieces n from each lot are to be tested and what is the sample average $\bar{X}$ which will differentiate acceptable from rejected lots?

We have

$$\bar{X}' = 35$$

$$\bar{X}'' = 30$$

$$P = 0.02$$

$$B = 0.01$$

The "population" of data meets the requirements for control laid down in the preceding chapter. To compute the standard deviation σ from the ranges: we know that for small samples all of size n and drawn from a normal population

$$\frac{\text{Mean range}}{\text{Standard deviation}} = \lambda$$

From our data and Table VI

$$\text{Mean range} = 14.0$$

$$\lambda = 2.970$$

$$\sigma = 4.714$$

From Table IV we find, corresponding to $P = 0.02$ and $B = 0.01$,

$$e = 2.054$$

$$f = 2.326$$

Finally

$$\frac{35-\bar{X}}{4.714/\sqrt{n}} = 2.054$$

$$\frac{\bar{X}-30}{4.714/\sqrt{n}} = 2.326$$

from which

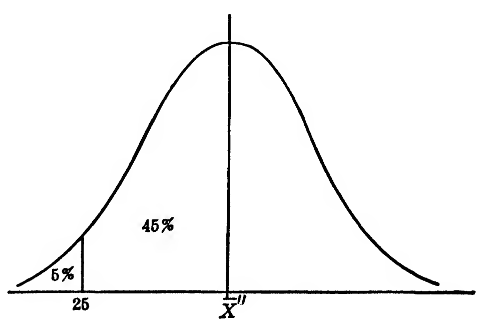
$$\bar{X} = 32.65$$

$$n = 17.1$$

i.e., a sample mean of 32.65 and a sample size of about 18.

Inasmuch as the lot is not indefinitely larger than the sample ($n = 18$) the results may be accepted only as rough approximations.

In place of the foregoing, the buyer may prefer to stipulate that he wishes to run, say, no more than 1 chance in 100 that he will receive a lot, a certain percentage of the items of which will be lower in quality than a stated value. From such a stipulation, we may easily compute $\bar{X}''$ and proceed as above. Thus if in the present example, the buyer wished to run only 1 chance in 100 of receiving lots with more than 5 per cent of their contents below a quality of 25, we would have what is shown graphically below.



From Table IV,

$$\frac{\bar{X}'' - 25}{4.714} = 1.65$$

or

$$\bar{X}'' = 32.778$$

This value of $\bar{X}''$ would now be placed in our equations and $\bar{X}$ and n could be found.

5.4 Producer and buyer risk, using fraction defective. As already illustrated, quality is sometimes not recorded numerically but simply as good or bad. We want to determine the size of sample n to be randomly drawn for inspection or test purposes from a lot of size N and also the maximum number of defective pieces g the sample may contain for the lot still to be acceptable. The producer, as before, is presumed to be manufacturing the product at a statistically controlled fraction defective p and he wishes to run a small risk P of having his lots rejected. The consumer wishes to run a small risk B of receiving lots which have more than a proportion of p' defective.

For these conditions, the producer's and consumer's interests are given, respectively, by

$$\sum_{r=g+1}^{r=n} \frac{C_r^{pN} C_{n-r}^{qN}}{C_n^N} = P$$

$$\sum_{r=0}^{r=g} \frac{C_r^{p'N} C_{n-r}^{q'N}}{C_n^N} = B$$

where

$$C_r^{pN} = \frac{(pN)!}{r!(pN-r)!}$$

with similar expressions for the other combinatorial terms. Considering the lot to be indefinitely large, these become

$$\sum_{r=g+1}^{r=n} C_r^n p^r q^{n-r} = P$$

$$\sum_{r=0}^{r=g} C_r^n p'^r q'^{n-r} = B$$

where

$$q = 1 - p, \quad q' = 1 - p'.$$

Finally, if p and p' are under 10 per cent, which is common in industrial practice, and n is large (say over 100) the Poisson form of the above equations may be safely used.

$$\sum_{r=g+1}^{r=n} \frac{e^{-pn}(pn)^r}{r!} = 1 - \sum_{r=0}^{r=g} \frac{e^{-pn}(pn)^r}{r!} = P$$

$$\sum_{r=0}^{r=g} \frac{e^{-p'n}(p'n)^r}{r!} = B$$

5.5 Example of 5.4. Supplement B of the American Society for Testing Materials' " Manual on Presentation of Data " (1) gives the following data on surface defects on galvanized hardware.

Lot number	Sample size	Number of defective pieces	Lot number	Sample size	Number of defective pieces
1	580	9	17	640	3
2	550	7	18	580	4
3	580	3	19	510	6
4	640	9	20	580	8
5	760	11	21	600	8
6	760	12	22	640	12
7	510	9	23	640	9
8	550	10	24	580	8
9	640	10	25	580	8
10	640	10	26	510	4
11	640	8	27	640	6
12	640	10	28	550	8
13	580	7	29	550	8
14	580	9	30	430	3
15	550	5	31	430	6
16	430	5	..	...	..

How many pieces should be taken in each sample, and what is the largest number of defective pieces a sample may contain for the lot still to be accepted?

The producer's mean quality is given by 0.013 and the population may be shown, by the methods of the preceding chapter, to be under statistical control. Assume that the producer wishes to run not more than 1 chance in 100 (because of high manufacturing costs) of having lots rejected while the consumer is willing to run as many as 5 chances in 100 (because of relative ease of replacement) of having as much as 5 per cent of the product defective. We have

$$p = 0.013$$

$$p' = 0.05$$

$$P = 0.01$$

$$B = 0.05$$

Assume an answer, $n = 200$. We have

$$\sum_{r=g+1}^{r=n} \frac{e^{-2.6}(2.6)^r}{r!} = 0.01$$

$$\sum_{r=g+1}^{r=n} \frac{e^{-10.0}(10.0)^r}{r!} = 1 - 0.05 = 0.95$$

From Figure 1, the first equation is satisfied by $g + 1 = 7.5$, approximately. But the second requires $g + 1 = 5.5$, so $n = 200$ is not a solution. By trial and error we come to $n = 300$ and $g + 1 = 9.7$ as an approximate solution. The sample size should be about 300 and the lot should not be accepted if it contains more than about nine defective pieces.

Surface defects may be easily noted and at slight inspection expense. In such cases, 100 per cent inspection might be feasible. This would not be true wherever inspection was costly or where destructive testing was necessary.

NOTES

5.6 Hypergeometric law. Given a well-mixed lot of size N with α bad pieces and β $(= N - \alpha)$ good pieces. A random sample of size n is drawn from the lot. What is the probability P that the sample contains a bad pieces and b good pieces?

A sample of size n can be drawn from a lot of size N in C_n^N ways. Further, a bad pieces can be drawn from α bad pieces in C_a^α ways. Similarly, b good pieces can be drawn in C_b^β ways. Each of C_a^α sets of a bad pieces can be paired with each of the C_b^β sets of b good pieces, i.e., the total number of ways in which a bad pieces and b good pieces can be drawn is $C_a^\alpha \cdot C_b^\beta$. Hence the required probability is

$$P = \frac{C_a^\alpha \cdot C_b^\beta}{C_n^N} = \frac{C_a^\alpha \cdot C_b^\beta}{C_{a+b}^{\alpha+\beta}}$$

which is sometimes known as the hypergeometric law.

Thus, if an urn contains two white and two black balls, the probability that a random sample of two consists of one white and one black ball is

$$\frac{C_1^2 \cdot C_1^2}{C_2^4} = \frac{4}{6}$$

Fallacious answers to problems of this type can be avoided if one enumerates the equally likely cases. Thus our lot is

A	B	C	D
0	0	●	●

and the following are equally likely drawings for a sample size of 2

A	B	B	C
0	0	0	●
A	C	B	D
0	●	0	●
A	D	C	D
0	●	●	●

four of which satisfy the requirements of the problem, i.e.,

$$P = \frac{4}{6}$$

The hypergeometric law may be looked upon as the law of compound probabilities for the case in which the several probabilities are affected by previous drawings. To illustrate, consider the following problem from Fry (17):

A batch of 1000 lamps is 5 per cent bad. If five are tested, what is the chance that no bad lamps will appear?

By the hypergeometric law

$$P = \frac{C_5^{950}\, C_0^{50}}{C_5^{1000}} = \frac{950!\, 995!}{945!\, 1000!} = 0.7734$$

The probability that the first lamp is good is 950/1000. If a good lamp is drawn *and not replaced*, the probability that the second lamp drawn is good is 949/999. Finally, as all five lamps must be good to satisfy the conditions of the problem,

$$P = \frac{950}{1000} \cdot \frac{949}{999} \cdot \frac{948}{998} \cdot \frac{947}{997} \cdot \frac{946}{996}$$

$$= \frac{950!}{1000!} \cdot \frac{995!}{945!} = 0.7734 \text{ as before}$$

5.7 Binomial approximation. If the lot size N is indefinitely larger than the sample size n, the probability that a lamp is bad will not vary as lamps are drawn. If $p = 0.05$ is the probability of drawing a bad lamp, the probability of drawing $0, 1, \cdots n$ good lamps in a sample of n is given by the successive terms of the binomial

$$(q + p)^n$$

Thus

$$P = (0.95)^5 = 0.7738$$

differing but slightly from the previous answer, as would be expected, for N is 200 times as large as n.

5.8 Poisson distribution. Given

$$(q + p)^n$$

write

$$P_r = C_r^n q^{n-r} p^r = \frac{n!}{r!\,(n-r)!}\, q^{n-r} p^r$$

P_r being the probability of obtaining exactly r defectives in a random sample of n pieces drawn from an indefinitely larger lot. If n is large, p small, and m (= pn, the expected number of defects) is a small finite number, the expression for P_r can be simplified. First write

$$q^{n-r} = (1-p)^{n-r} = \left(1 - \frac{m}{n}\right)^n \left(1 - \frac{m}{n}\right)^{-r}$$

For r considerably less than n, the last factor will not differ appreciably from unity, i.e.,

$$q^{n-r} \approx q^n$$

Replacing $n!$ and $(n-r)!$ by the Stirling approximations

$$n! = \sqrt{2\pi n}\, n^n\, e^{-n}$$

$$(n-r)! = \sqrt{2\pi(n-r)}\, (n-r)^{n-r} e^{-(n-r)}$$

we obtain

$$P_r = \frac{n^r}{r!}(1-p)^n\, p^r$$

$$= \frac{m^r}{r!}\left(1 - \frac{m}{n}\right)^n$$

$$= \frac{m^r e^{-m}}{r!}$$

the equation of the Poisson distribution.

In our first example $n = 5$, $p = 1/20$, $pn = 1/4$. The conditions for proper application of the Poisson approximation are not satisfied, for n is not large. The result indicates, however, a close approximation to the exact answer as given by the hypergeometric law, for

$$P_0 = \frac{m^0 e^{-1/4}}{0!} = 0.7788$$

5.9 Note on Figure 1 (p. 176). In a sample of size n, the probabilities of 0, 1, $\cdots$, n defects, as given by the Poisson law, are

$$\frac{e^{-m}}{0!}, \frac{e^{-m}m}{1!}, \frac{e^{-m}m^2}{2!}, \ldots, \frac{e^{-m}m^n}{n!}$$

The probability that a sample of n contains more than g defects (i.e., at least $g + 1$ defects) is

$$\sum_{r=g+1}^{r=n} \frac{e^{-m}m^r}{r!} .$$

Figure 1 gives the probability of at least c defects for various values of m, that is,

$$\sum_{r=c}^{r=n} \frac{e^{-m}m^r}{r!}$$

which is equivalent to

$$1 - \sum_{r=0}^{r=c-1} \frac{e^{-m}m^r}{r!}$$

5.10 Producer and buyer risk, using the standard deviation. Frequently, variability in the quality of a product may be even less desirable than low average quality. Metal strips all of about the same breaking strength and electric lamps all of about the same life may often be preferred to batches of these products which are of higher mean quality but which contain many very good and many very bad strips or lamps.

Crum (7) states that studies involving several hundred concrete beams used in paving projects in Iowa yield a standard deviation σ' of about 10 per cent of the mean quality. The latter is given by a modulus of rupture of 760 pounds per square inch. Assume that for a certain job, $\sigma'' = 20\%$ is the buyer's tolerance variability. Producer and buyer want small risks, say, 1 in 100 of respectively (1) rejected lots, (2) less-than-tolerance quality lots. How many pieces should be drawn for test purposes from each lot and what should be the maximum standard deviation of the sample if that lot is to be accepted?

The distribution of the variances s_i^2 of samples each of size n drawn at random from a normal population of variance σ^2 is given by

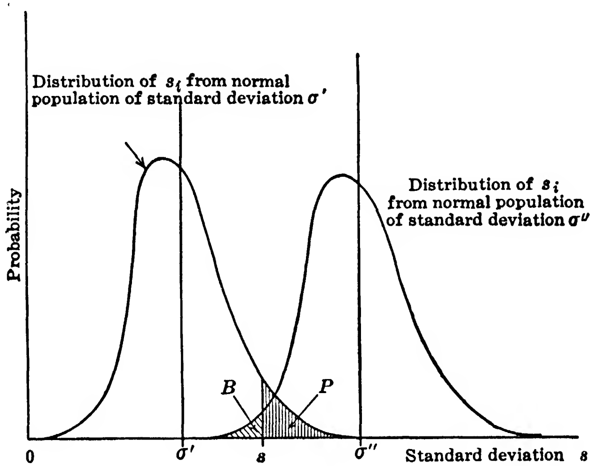
$$p(s^2)\, d(s^2) = C(s^2)^{n-3/2}\, e^{-ns^2/2\sigma^2}\, d(s^2)$$

where C is a constant. The distribution of s_i is immediately derivable and has been discussed in 1.37. It is most convenient, so far as available tables are concerned, to use the fact that the function

$$\frac{ns^2}{\sigma^2}$$

is distributed as χ^2 with $n - 1$ degrees of freedom. Values of χ^2 are shown in Table VII.

In this example, the larger the value of σ, the poorer the quality. Hence, compared to the two previous examples, the producer's and the tolerance populations are reversed in position along the horizontal axis.



We have

$$P \text{ (producer's risk)} = 0.01$$

$$B \text{ (buyer's risk)} = 0.01$$

$$\sigma' = 76 \qquad \sigma'' = 152$$

$$\sigma'^2 = 5776 \qquad \sigma''^2 = 23{,}104$$

We want to determine n and s. We have for the population of $\sigma' = 76$

[1] $\dfrac{ns^2}{5776} = \chi_P^2$ with $n - 1$ degrees of freedom, and for the population of tolerance quality $\sigma'' = 152$

[2] $\dfrac{ns^2}{23{,}104} = \chi_B^2$ with $n - 1$ degrees of freedom

In Table VII we have the probabilities of *exceeding* a value of χ^2 for various degrees of freedom. We know neither χ^2 nor the number of degrees of freedom, but we do know that we want the producer's chance of exceeding χ_P^2 to be 0.01 and the buyer's chance of exceeding χ_B^2 to be 0.99. Also, from [1] and [2]

$$\frac{\chi_P^2}{\chi_B^2} = 4$$

From Table VII, using columns headed by probabilities of 0.99 and 0.01, we find for the above ratio of 4 about 24 degrees of freedom. (χ_P^2 = 42.980, χ_B^2 = 10.856.) Hence

$$n = 25$$

Substituting in either [1] or [2] we find

$$s^2 = 9930, \text{ approximately}$$
$$s = 99.7$$

A sample should contain 25 items and have a standard deviation of not more than 100 pounds per square inch for its lot to be acceptable.

5.11 Second example of 5.10. Welch (46) has given examples in which the size of the sample has, as is sometimes the case, already been fixed by authority. A manufacturer produces electric light bulbs under controlled conditions with $\sigma' = 0.8$. Ten bulbs are to be sampled from each lot. The producer is willing to incur a 5 per cent risk of having lots rejected whereas the buyer wants to know what protection such a sample will, under these conditions, give him against obtaining lots as bad as $\sigma'' = 1.5$.

We have

$$\frac{ns^2}{\sigma'^2} = 15.62s^2 = \chi_P^2$$

For a producer risk of 0.05 and for nine degrees of freedom we have $\chi_P^2 = 16.919$. Hence

$$s = 1.04$$

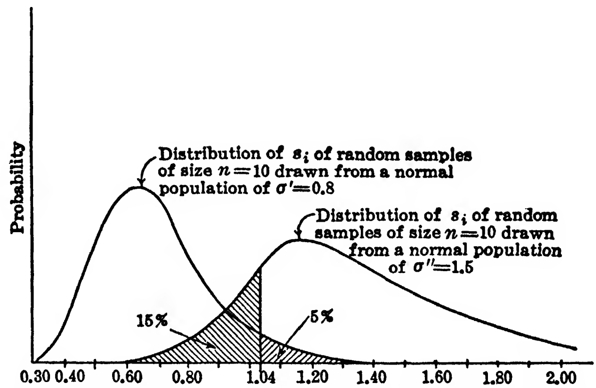
Finally

$$\frac{ns^2}{\sigma''^2} = \frac{10(1.04)^2}{(1.5)^2} = 4.81 = \chi_B^2$$

For $\chi_B^2 = 4.85$ and again with nine degrees of freedom, we have $P = 0.85$, which is the chance of exceeding χ_B^2. The buyer's risk is therefore 1.00 − 0.85, or 15 per cent, a rather high risk. For better protection to him either the sample size n must be increased or the standard deviation σ' reduced by more efficient plant control.*

* If the producer uses inspection for control, then it is easily shown that for $\sigma = 0.8$ and $n = 10$, a value of $s = 1.04$ will likely result in the lot being thrown out; if so, the buyer's risk is practically zero. This applies in principal to all problems in this chapter. We presume, however, that the buyer desires protection quite independent of the producer's intentions.

The following graph illustrates the conditions and conclusions of the preceding example.



General character of distribution of standard deviations s_i of samples of size $n = 10$

5.12 Producer and buyer risk, using the coefficient of variation. Specification of average quality and variability in quality may be separately provided by the methods already discussed. It is sometimes desired to make use of one hybrid statistic which has both features. One is the coefficient of variation which is given by

$$\frac{\text{Standard deviation}}{\text{Arithmetic mean}}$$

High values of this statistic will result from high variability in quality and low mean quality, both of which we take in our examples to be unfavorable. Correspondingly, low values of the coefficient of variation are considered favorable.

Wilsdon (49) gives a frequency distribution showing the crushing strength, in tons, of 188 tests of a brand of brick. As the original data are not shown, the following population parameters are estimated from his frequency distribution of 188 observations.

$$\bar{X}' = 28.1$$

$$\sigma' = 4.1$$

or the coefficient of variation at the works is $V_P = 0.1459$.

Assume that for a certain purpose a buyer is willing to accept brick of lower average strength and higher variability in strength. He is willing to run a risk of 5 chances in 100 (B) of receiving lots of coefficient of variation $V_B = 0.3$. The producer whose statistically controlled output

is presumed to be characterized by $V_P = 0.1459$ wishes to run, say, no greater than a 1 in 100 risk (P) of having lots rejected. How many bricks should be tested, and what is the sample coefficient of variation v which divides acceptable from unacceptable lots?

The function

$$\frac{nv^2}{1 + v^2}\left(\frac{1}{V^2} + 1\right)$$

is distributed approximately as χ^2 with $n - 1$ degrees of freedom. This approximation is sufficiently accurate for $V \not> 1/3$ and $n \not< 6$.

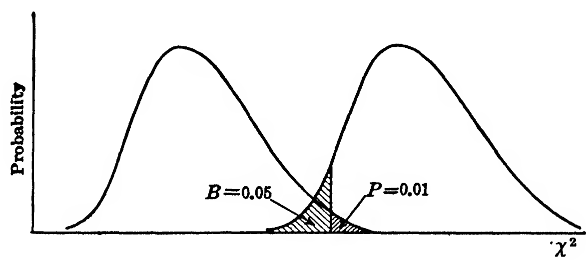
We have, for producer and buyer interests respectively,

$$\frac{nv^2}{1 + v^2}\left(\frac{1}{V_P^2} + 1\right) = \chi_P^2$$

$$\frac{nv^2}{1 + v^2}\left(\frac{1}{V_B^2} + 1\right) = \chi_B^2$$

where $V_P = 0.1459$ is the coefficient of variation associated with the producer's ordinary output and $V_B = 0.3$ is the coefficient of variation associated with the buyer's tolerance output. Correspondingly, χ_P^2 is the value of χ^2 associated with the producer's risk (0.01) and χ_B^2 the value of χ^2 associated with the buyer's risk (0.05); to find v and n.

The following graphical description may be useful.



As before, we divide one equality by the other and obtain

$$\frac{\chi_P^2}{\chi_B^2} = \frac{48}{12} = 4$$

Entering the χ^2 tables with probabilities of 0.01 and 0.95, we find the ratio 4 to be associated with approximately 16 degrees of freedom. Hence

$$n = 17$$

Substituting this value of n and the appropriate value of χ^2 into either

of the two original equations, we find

$$v = 0.202$$

A sample of $n = 17$ should be drawn and a sample coefficient of $v = 0.202$ should divide acceptable and non-acceptable lots.

5.13 Normal approximation to χ^2. If instead of $V_B = 0.3$, we had a more stringent buyer's tolerance level, say, $V_B = 0.2$, we would have found

[3]
$$\frac{\chi_P^2}{\chi_B^2} = 1.83$$

For this ratio no satisfying number of degrees of freedom can be found in the tables of χ^2, i.e., $n - 1$ is greater than 30. In such a case, a normal distribution solution is possible, for

$$\sqrt{2\chi^2} - \sqrt{2n - 3}$$

is distributed normally with unit variance.

We have

[4]
$$\sqrt{2\chi_P^2} - \sqrt{2n - 3} = 2.32$$
$$\sqrt{2\chi_B^2} - \sqrt{2n - 3} = -1.65$$

the values 2.32 and -1.65 (associated with producer and buyer risks of 0.01 and 0.05 respectively) being found from Table IV. From [3] and [4] we obtain

$$n = 85, \text{ approximately}$$
$$v = 0.172$$

As would be expected, if the buyer is to be protected against quality lower than $V_B = 0.2$ (instead of $V_B = 0.3$) a larger sample and a smaller sample coefficient of variation will be required.

5.14 Second Example of 5.12. Examples may be given in which n has already been specified by an industrial agreement or by a governmental authority. Pearson (31, *b*) considers a case in which $V_B = 0.200$ and $n = 12$. At what level V_P must the producer control the quality of his product, and what shall be the value of v which separates acceptable from non-acceptable lots, in order that the buyer shall run a 1 per cent chance (B) of obtaining lots whose quality is given by a coefficient of variation $V_B = 0.200$, and the producer a 5 per cent chance (P) of having lots rejected? We have

$$\chi_B^2 = 312\frac{v^2}{1 + v^2}$$

For 11 degrees of freedom and for $B = 0.01$ (which is equivalent to 99 chances in 100 of exceeding χ_B^2) we have from Table VII,

$$\chi_B^2 = 3.053$$

from which

$$v = 0.0995$$

To calculate the necessary level of control V_P,

$$\frac{nv^2}{1 + v^2}\left(\frac{1}{V_P^2} + 1\right) = \chi_P^2$$

we have

$$n = 12$$

$$v = 0.0995$$

$$\chi_P^2 = 19.675$$

from which

$$V_P = 0.0776$$

NOTE

5.15 Distribution of *v*. McKay (27) is responsible for the proof that

$$\frac{nv^2}{1 + v^2}\left(\frac{1}{V^2} + 1\right)$$

is distributed approximately as χ^2 with $n - 1$ degrees of freedom. The approximation is best when the coefficient of variation (V) of the normal population is small. Fieller (14) gives the following numerical results which show that as n becomes larger, the χ^2 approximation improves.

$V = \frac{1}{3}, n = 6$

Chance of sample with smaller v			Chance of sample with larger v		
v	True value	χ^2 theory	v	True value	χ^2 theory
0.15	0.062	0.067	0.48	0.053	0.048
0.14	0.047	0.051	0.51	0.034	0.030
0.13	0.034	0.037	0.54	0.022	0.019
0.12	0.024	0.026	0.57	0.013	0.012
0.11	0.017	0.018	0.60	0.008	0.007
0.10	0.011	0.012	0.63	0.005	0.004
0.09	0.007	0.007	0.66	0.003	0.003
0.08	0.004	0.004	0.69	0.002	0.002
0.07	0.002	0.002	0.72	0.001	0.001
0.06	0.001	0.001	0.75	0.001	0.001

$V = \frac{1}{3}, n = 18$

Chance of sample with smaller v			Chance of sample with larger v		
v	True value	χ^2 theory	v	True value	χ^2 theory
0.24	0.084	0.088	0.42	0.060	0.058
0.23	0.058	0.061	0.43	0.046	0.044
0.22	0.038	0.040	0.44	0.035−	0.033
0.21	0.024	0.026	0.45	0.026	0.024
0.20	0.014	0.015+	0.46	0.019	0.018
0.19	0.008	0.009	0.47	0.014	0.013
0.18	0.004	0.005−	0.48	0.010	0.009
0.17	0.002	0.002	0.49	0.007	0.007
0.16	0.001	0.001	0.50	0.005−	0.005−
....	.....	.....	0.51	0.003	0.003
....	.....	.....	0.52	0.002	0.002
....	.....	.....	0.53	0.002	0.002
....	.....	.....	0.54	0.001	0.001
....	.....	.....	0.55	0.001	0.001

ACKNOWLEDGMENT FOR DATA

I wish to thank all who kindly permitted me to use their data. In several instances, particularly with data from foreign publications, permission has not been requested. I hope the following citations will be accepted as acknowledgment of my great obligation to all whose data were used.

REFERENCES

1. American Society for Testing Materials, *Supplement B, Manual on Presentation of Data*, 1933.
2. Beckwith, O., *Report to the Product Engineering Department of Alexander Smith and Sons Carpet Company*, November 7, 1938.
3. Bowley, A. L., *Elements of Statistics*, 6th edition. London, P. S. King and Son.
4. Brenner, A., *Magnetic Method for Measuring the Thickness of Non-Magnetic Coatings on Iron and Steel*, National Bureau of Standards, RP 1081, March, 1938.
5. British Cotton Industry Research Association, *An Explanation of Some Statistical Methods Used in Textile Research*, Shirley Institute, Didsbury, Manchester, January, 1929.
6. Campbell, J. M., and W. G. Lovell, "Application of Statistical Concepts to the Knock-Rating Problem," *Society of Automotive Engineers' Journal*, Vol. 43, No. 4, October, 1938.
7. Crum, R. W., "The Numbers of Specimens or Tests of Concrete and Concrete Aggregates Required for Reasonable Accuracy of the Average," Report on Significance of Tests of Concrete and Concrete Aggregates, *American Society for Testing Materials*, 1935.
8. Deming, W. Edwards, and Raymond T. Birge, "On the Statistical Theory of Errors," The Graduate School of the U. S. Department of Agriculture, Additional Notes 1937, 1938.
9. Dodge, H. F., "Acceptance-Rejection Requirements in Specifications," *Proceedings of the American Society for Testing Materials*, Vol. 34, Part II, 1934.
10. Dudding, B. P., and I. M. Baker, "The Application of Statistical Methods to the Planning of Routine Testing Procedure," *Transactions of the Society of Glass Technology*, Vol. 18, 1934.
11. Editorial, "Fundamentals of the Theory of Sampling," *Annals of Mathematical Statistics*, Vol. 1, No. 1, February, 1930.
12. Eisenhart, C., "The Interpretation of Certain Regression Methods and Their Use in Biological and Industrial Research," *Annals of Mathematical Statistics*, Vol. X, No. 2, June, 1939.
13. Fieldner, A. C., and W. A. Selvig, "Notes on the Sampling and Analysis of Coal," Technical Paper 586, Bureau of Mines, Washington, 1938.
14. Fieller, E. C., "A Numerical Test of the Adequacy of A. T. McKay's Approximation," *Journal of the Royal Statistical Society*, Part IV, 1932.

15. FISHER, R. A., (*a*) "Application of Student's Distribution," *Metron*, Vol. 5, 1925; (*b*) "The Moments of the Distribution for Normal Samples of Measures of departure from Normality," *Proceedings of the Royal Society, A*, Vol. CXXX.
16. FISHER, R. A., and F. YATES, *Statistical Tables for Biological, Agricultural and Medical Research*, Edinburgh, Oliver and Boyd, 1938.
17. FRY, T. C., *Probability and Its Engineering Uses*, New York, D. Van Nostrand Company, 1928.
18. FULWEILER, W. H., A. H. STANG, and L. R. SWEETMAN, "Inspection and Tensile Tests of Some Worn Wire Ropes," National Bureau of Standards, RP 920, Vol. 17, September, 1936.
19. FURRY, MARGARET S., "Breaking Strength, Elongation, and Folding Endurance of Films of Starches and Gelatin Used in Textile Sizing" (Statistical analysis by G. W. Snedecor and Gertrude M. Cox), U. S. Department of Agriculture, Technical Bulletin 674, April, 1939.
20. GEARY, R. C., and E. S. PEARSON, "Tests of Normality," Issued by the *Biometrika* Office, University College, London, University Press, Cambridge.
21. GOULD, C. E., and W. M. HAMPTON, "Statistical Methods Applied to the Manufacture of Spectacle Glasses," *Supplement to the Journal of the Royal Statistical Society*, Vol. III, No. 2, 1936.
22. IRWIN, J. O., "Mathematical Theorems Involved in the Analysis of Variance," *Journal of the Royal Statistical Society*, Vol. 94, 1931.
23. JACKSON, DUNHAM, "The Theory of Small Samples," *American Mathematical Monthly*, Vol. 42, No. 6.
24. JENNETT, W. J., and B. P. DUDDING, "The Application of Statistical Principles to an Industrial Problem," *Supplement to the Journal of the Royal Statistical Society*, Vol. III, No. 1, 1936.
25. LOGAN, K. H., and S. P. EWING, "Soil Corrosion Studies, 1934; Field Tests of Non-Bituminous Coatings for Underground Use," U. S. Bureau of Standards, Vol. 18, March, 1937.
26. MAHALANOBIS, P. C., "Tables for L Tests," *Sankhyā (The Indian Journal of Statistics)*, Vol. I, Part 1, June, 1933.
27. MCKAY, A. T., "Distribution of the Coefficient of Variation and the Extended t-Distribution," *Journal of the Royal Statistical Society*, Part IV, 1932.
28. NAYER, P. P. N., "An Investigation into the Application of Neyman and Pearson's L_1 Test with Tables of Percentage Limits," *Statistical Research Memoirs*, Vol. 1, London.
29. NESS, H. C., "Sampling When the Parent Population is Pearson Type X," M. I. T., 1938.
30. NEYMAN, J., and E. S. PEARSON, "On the Problem of k Samples," *Bulletin de l'Acadamie Polonaise des Sciences et des Lettres*, Serie A, 1931.
31. PEARSON, E. S., (*a*) "A Further Development of Tests for Normality," *Biometrika*, Vol. xxii, p. 239; (*b*) "Statistical Method in the Control and Standardization of the Quality of Manufactured Products," *Journal of the Royal Statistical Society*, Vol. XCVI, Part 1, 1933.
32. PETTEBONE, E. R. II, and L. C. YOUNG, "Application of Statistical Methods to the Control of Quality of Manufactured Gas," M. I. T. Thesis, 1937.
33. PULSIFER, H. B., "Physical Properties of Fine Bolts," *Automotive Industry*, Vol. 63, 1930.
34. RIDER, P., (*a*) *An Introduction to Modern Statistical Methods*, New York, John Wiley and Sons, 1939, Western Electric Company data with the kind permission

of that company and Mr. A. G. Brooks; (*b*) " A Survey of the Theory of Small Samples," *Annals of Mathematics*, Second Series, Vol. 31, No. 4, October, 1930.

35. Rietz, H. L., " Some Topics in Sampling Theory," *Bulletin of the American Mathematical Society*, Vol. 43, 1937.
36. Shewhart, W. A., *Economic Control of the Quality of Manufactured Product*, New York, D. Van Nostrand Company, 1931.
37. Shoumatoff, N., " An Application of Statistical Methods to Quality Control in a Radio Tube Plant," M. I. T. Thesis, 1939.
38. Snedecor, George W., " Calculation and Interpretation of the Analysis of Variance and Covariance," Collegiate Press, Ames, Iowa, 1934.
39. " Student," " The Probable Error of the Mean," *Biometrika*, Vol. VI, 1908.
40. Sukhatme, P. V., *On Fisher and Behren's Test of Significance for the Difference in Means of Two Normal Samples*, Proceedings of the First Session of the Indian Statistical Conference held in Calcutta, January, 1938; December, 1938, Statistical Publishing Society, Calcutta.
41. Thomsen, P., *Outline of a Course in Statistical Methods for Chemists and Engineers*, Procter and Gamble Company.
42. Tippett, L. H. C., (*a*) *The Methods of Statistics*, London, Williams and Norgate, Ltd., 2d ed., 1937; (*b*) *The Statistical Principles of Experimentation, with Particular Reference to Experimentation in Textile Factories*, Proceedings of the Industrial Statistics Conference held at M. I. T.; New York, Pitman Publishing Corporation, 1939.
43. Van Rest, E. D., " Examples of Statistical Methods in Forest Products Research," *Supplement to the Journal of the Royal Statistical Society*, Vol. IV, No. 2, 1937.
44. Wald, Abraham, " The Fitting of Straight Lines if Both Variables are Subject to Error," *Annals of Mathematical Statistics*, Vol. XI, No. 3, September, 1940.
45. Wald, A., and J. Wolfowitz, " On a Test Whether Two Samples are from the Same Population," *Annals of Mathematical Statistics*, Vol. XI, No. 2, June, 1940.
46. Welch, B. L., " Specification of Rules for Rejecting Too Variable a Product with Particular Reference to an Electric Lamp," *Supplement to the Journal of the Royal Statistical Society*, Vol. III, No. 1, 1936.
47. Whittaker, E. T., and G. Robinson, *The Calculus of Observations*, London, Blackie and Son, Limited, 1937.
48. Wilks, S. S., *Statistical Inference 1936–1937*, Ann Arbor, Michigan, Edwards Brothers.
49. Wilsdon, B. H., " Standard Specifications for Clay Building Materials," *The British Clayworker*, May, 1932.

GENERAL REFERENCES ON METHODS

Several of the following books should be consulted in connection with the material of this book.

Aitken, A. C., *Statistical Mathematics*, Edinburgh, Oliver and Boyd, 1939.

An introduction to mathematical statistics; considerably more elementary than Wilks' monograph.

Fisher, R. A., *Statistical Methods for Research Workers*, 6th ed., Edinburgh, Oliver and Boyd, 1936.

A book in which the author presents a unified treatment of modern experimental statistics, much of which is the result of his own contributions.

FISHER, R. A., *The Design of Experiments*, Edinburgh, Oliver and Boyd, 1937.

The leading discussion of the place of statistical technique in the planning of field and laboratory experiments. The author is the originator of many of the proposed methods.

MAINLAND, DONALD, *The Treatment of Clinical and Laboratory Data*, Edinburgh, Oliver and Boyd, 1938.

Although the author's examples are from medical and dental research, the exposition of the assumptions and background of statistics and the treatment of statistical difficulties encountered by research workers are not to be found in other studies.

PEARSON, E. S., *Application of Statistical Methods to Industrial Standardisation and Quality Control*, London, British Standards Institution, 1935.

An introduction to quality control for industrial workers who have had practically no previous training in statistics.

RIDER, P. R., *An Introduction to Modern Statistical Methods*, New York, John Wiley and Sons, 1939.

A textbook on modern statistical methods.

SHEWHART, W. A., *Economic Control of the Quality of Manufactured Product*, New York, D. Van Nostrand Company, 1931.

A general treatise on the subject of statistical methods for quality control.

SIMON, LESLIE E., *Engineers' Manual of Statistical Methods*, New York, John Wiley and Sons, 1941.

This book is intended for those engaged in systematic quality control. There are numerous examples of the author's work in the United States Army.

SNEDECOR, G. W., *Statistical Methods Applied to Experiments in Agriculture and Biology*, Ames, Iowa, Collegiate Press, 1938.

A textbook, written with appreciation of beginners' difficulties, with many applications from the experience of the author and his associates.

TIPPETT, L. H. C., *The Methods of Statistics*, 2d ed., London, Williams and Norgate, Ltd., 1937.

A book for experimentalists, with many practical examples.

WILKS, S. S., *Statistical Inference, 1936–1937*, Ann Arbor, Michigan, Edwards Brothers.

For those with adequate training in mathematics who wish to get to the mathematical foundations of modern statistical technique.

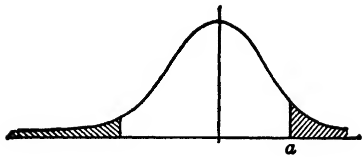


TABLE III

PROBABILITY POINTS OF a

Size of sample n	Probability points						Mean
	Upper 1%	Upper 5%	Upper 10%	Lower 10%	Lower 5%	Lower 1%	
11	0.9359	0.9073	0.8899	0.7409	0.7153	0.6675	0.81805
16	0.9137	0.8884	0.8733	0.7452	0.7236	0.6829	0.81128
21	0.9001	0.8768	0.8631	0.7495	0.7304	0.6950	0.80792
26	0.8901	0.8686	0.8570	0.7530	0.7360	0.7040	0.80590
31	0.8827	0.8625	0.8511	0.7559	0.7404	0.7110	0.80456
36	0.8769	0.8578	0.8468	0.7583	0.7440	0.7167	0.80360
41	0.8722	0.8540	0.8436	0.7604	0.7470	0.7216	0.80289
46	0.8682	0.8508	0.8409	0.7621	0.7496	0.7256	0.80233
51	0.8648	0.8481	0.8385	0.7636	0.7518	0.7291	0.80188
61	0.8592	0.8434	0.8349	0.7662	0.7554	0.7347	0.80122
71	0.8549	0.8403	0.8321	0.7683	0.7583	0.7393	0.80074
81	0.8515	0.8376	0.8298	0.7700	0.7607	0.7430	0.80038
91	0.8484	0.8353	0.8279	0.7714	0.7626	0.7460	0.80010
101	0.8460	0.8344	0.8264	0.7726	0.7644	0.7487	0.79988
201	0.8322	0.8229	0.8178	0.7796	0.7738	0.7629	0.79888
301	0.8260	0.8183	0.8140	0.7828	0.7781	0.7693	0.79855
401	0.8223	0.8155	0.8118	0.7847	0.7807	0.7731	0.79838
501	0.8198	0.8136	0.8103	0.7861	0.7825	0.7757	0.79828
601	0.8179	0.8123	0.8092	0.7873	0.7838	0.7776	0.79822
701	0.8164	0.8112	0.8084	0.7878	0.7848	0.7791	0.79817
801	0.8152	0.8103	0.8077	0.7885	0.7857	0.7803	0.79813
901	0.8142	0.8096	0.8071	0.7890	0.7864	0.7814	0.79811
1001	0.8134	0.8090	0.8066	0.7894	0.7869	0.7822	0.79808

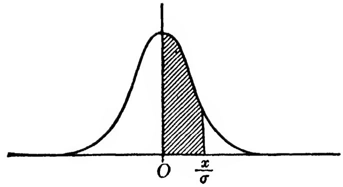


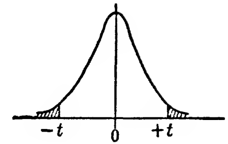
TABLE IV

Normal Distribution Areas

$\frac{x}{\sigma}$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0159	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0 0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2518	0.2549
0.7	0.2580	0.2612	0.2642	0.2673	0.2704	0.2734	0 2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3718	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4083	0 4099	0 4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4430	0.4441
1.6	0.4452	0.4463	0 4474	0.4485	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4758	0.4762	0.4767
2.0	0.4773	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4865	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4980	0.4980	0.4981
2.9	0.4981	0.4982	0.4983	0.4984	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.49865	0.4987	0.4987	0.4988	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990
3.1	0.49903	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993129									
3.3	0.4995166									
3.4	0.4996631									
3.5	0.4997674									
3.6	0.4998409									
3.7	0.4998922									
3.8	0.4999277									
3.9	0.4999519									
4.0	0.4999683									
4.5	0.4999966									
5.0	0.4999997133									

TABLE V

t



Degrees of freedom	P = 0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.05	0.02	0.01
1	0.158	0.325	0.510	0.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	0.142	0.289	0.445	0.617	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.137	0.277	0.424	0.584	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.134	0.271	0.414	0.569	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.132	0.267	0.408	0.559	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.131	0.265	0.404	0.553	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.130	0.263	0.402	0.549	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.130	0.262	0.399	0.546	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.129	0.261	0.398	0.543	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250
10	0.129	0.260	0.397	0.542	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169
11	0.129	0.260	0.396	0.540	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106
12	0.128	0.259	0.395	0.539	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055
13	0.128	0.259	0.394	0.538	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012
14	0.128	0.258	0.393	0.537	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977
15	0.128	0.258	0.393	0.536	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947
16	0.128	0.258	0.392	0.535	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921
17	0.128	0.257	0.392	0.534	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898
18	0.127	0.257	0.392	0.534	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878
19	0.127	0.257	0.391	0.533	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861
20	0.127	0.257	0.391	0.533	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845
21	0.127	0.257	0.391	0.532	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831
22	0.127	0.256	0.390	0.532	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819
23	0.127	0.256	0.390	0.532	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807
24	0.127	0.256	0.390	0.531	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797
25	0.127	0.256	0.390	0.531	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787
26	0.127	0.256	0.390	0.531	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779
27	0.127	0.256	0.389	0.521	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771
28	0.127	0.256	0.389	0.530	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763
29	0.127	0.256	0.389	0.530	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756
30	0.127	0.256	0.389	0.530	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750
∞	0.12566	0.25335	0.38532	0.52440	0.67449	0.84162	1.03643	1.28155	1.64485	1.95996	2.32634	2.57582

Reproduced from *Statistical Methods for Research Workers*, 6th ed., with the permission of the author, R. A. Fisher, and his publisher, Oliver and Boyd, Edinburgh.

TABLE VI

Ratio of the Mean Range to Standard Deviation

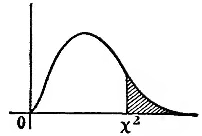
The ratio of mean range of samples of size n to σ of the normal population from which they are drawn

n	$\frac{\text{Mean range}}{\sigma}$	n	$\frac{\text{Mean range}}{\sigma}$	n	$\frac{\text{Mean range}}{\sigma}$	n	$\frac{\text{Mean range}}{\sigma}$
0	....	10	3.07751	20	3.73495	30	4.08552
1	....	11	3.17287	21	3.77834	31	4.11293
2	1.12838	12	3.25846	22	3.81938	32	4.13934
3	1.69257	13	3.33598	23	3.85832	33	4.16482
4	2.05875	14	3.40676	24	3.89535	34	4.18943
5	2.32593	15	3.47183	25	3.93063	35	4.21322
6	2.53441	16	3.53198	26	3.96432	36	4.23625
7	2.70436	17	3.58788	27	3.99654	37	4.25855
8	2.84720	18	3.64006	28	4.02741	38	4.28018
9	2.97003	19	3.68896	29	4 05704	39	4.30117

n	$\frac{\text{Mean range}}{\sigma}$	n	$\frac{\text{Mean range}}{\sigma}$	n	$\frac{\text{Mean range}}{\sigma}$	n	$\frac{\text{Mean range}}{\sigma}$
40	4.32156	85	4.89789	150	5.29849	400	5.93636
45	4.41544	90	4.93940	160	5.34244	450	6.00903
50	4.49815	95	4.97841	170	5.38344	500	6.07340
55	4.57197	100	5.01519	180	5.42186	600	6.18340
60	4.63856	105	5.04997	190	5.45799	700	6.27510
65	4.69916	110	5.08295	200	5.49209	800	6.35358
70	4.75472	120	5.14417	250	5.63837	900	6.42211
75	4.80598	130	5.19996	300	5.75553	1000	6.48287
80	4.85355	140	5.25118	350	5.85302	....	....

TABLE VII

χ^2



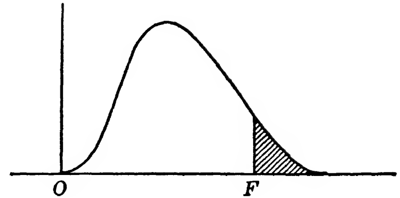
Degrees of freedom	P = 0.99	0.98	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.02	0.01
1	0.000157	0.000628	0.00393	0 0158	0.0642	0.148	0.455	1.074	1.642	2.706	3.841	5.412	6.635
2	0.0201	0.0404	0.103	0.211	0 446	0.713	1.386	2.408	3.219	4.605	5.991	7.824	9.210
3	0.115	0.185	0.352	0.584	1.005	1.424	2.366	3.665	4.642	6.251	7.815	9.837	11.341
4	0.297	0.429	0.711	1.064	1.649	2.195	3.357	4.878	5.989	7.779	9.488	11.668	13.277
5	0.554	0.752	1.145	1.610	2.343	3.000	4.351	6.064	7.289	9.236	11.070	13.388	15.086
6	0.872	1.134	1.635	2.204	3.070	3.828	5.348	7.231	8.558	10.645	12.592	15.033	16.812
7	1.239	1.564	2.167	2.833	3.822	4.671	6.346	8.383	9.803	12.017	14.067	16.622	18.475
8	1.646	2.032	2.733	3.490	4.594	5 527	7.344	9.524	11.030	13.362	15.507	18.168	20.090
9	2.088	2.532	3.325	4.168	5.380	6.393	8.343	10.656	12.242	14.684	16.919	19.679	21.666
10	2.558	3.059	3.940	4.865	6.179	7.267	9.342	11.781	13.442	15.987	18.307	21.161	23.209
11	3.053	3.609	4.575	5.578	6.989	8.148	10.341	12.899	14.631	17.275	19.675	22.618	24.725
12	3.571	4.178	5.226	6.304	7.807	9.034	11.340	14 011	15.812	18.549	21.026	24.054	26.217
13	4.107	4.765	5.892	7.042	8.634	9 926	12.340	15.119	16.985	19.812	22.362	25.472	27.688
14	4.660	5.368	6.571	7.790	9.467	10 821	13.339	16.222	18.151	21.064	23.685	26.873	29.141
15	5.229	5.985	7.261	8 547	10.307	11.721	14.339	17.322	19.311	22.307	24.996	28.259	30.578
16	5.812	6.614	7.962	9 312	11.152	12.624	15.338	18.418	20.465	23.542	26.296	29.633	32.000
17	6.408	7.255	8.672	10 085	12.002	13 531	16.338	19.511	21.615	24.769	27.587	30.995	33.409
18	7.015	7.906	9.390	10 865	12.857	14.440	17.338	20.601	22.760	25 989	28.869	32.346	34.805
19	7.633	8 567	10.117	11 651	13.716	15 352	18.338	21.689	23.900	27.204	30.144	33.687	36.191
20	8.260	9.237	10.851	12.443	14.578	16.266	19.337	22.775	25.038	28.412	31.410	35.020	37.566
21	8.897	9.915	11.591	13.240	15.445	17.182	20.337	23.858	26.171	29.615	32.671	36.343	38.932
22	9.542	10.600	12.338	14.041	16.314	18.101	21.337	24.939	27.301	30.813	33.924	37.659	40.289
23	10.196	11.293	13 091	14 848	17 187	19.021	22.337	26 018	28.429	32.007	35.172	38.968	41.638
24	10.856	11.992	13.848	15.659	18 062	19.943	23.337	27.096	29.553	33.196	36.415	40.270	42.980
25	11.524	12 697	14.611	16.473	18.940	20.867	24.337	28.172	30.675	34.382	37.652	41.566	44.314
26	12.198	13.409	15.379	17 292	19.820	21.792	25.336	29.246	31.795	35.563	38.885	42.856	45.642
27	12.879	14.125	16.151	18.114	20.703	22.719	26.336	30.319	32.912	36.741	40.113	44.140	46.963
28	13 565	14.847	16.928	18.939	21.588	23.647	27.336	31.391	34.027	37.916	41.337	45.419	48.278
29	14.256	15.574	17.708	19.768	22.475	24.577	28.336	32.461	35.139	39.087	42.557	46.693	49.588
30	14.953	16.306	18.493	20.599	23.364	25.508	29.336	33.530	36.250	40.256	43.773	47.962	50.892

For degrees of freedom greater than 30, the expression $\sqrt{2\chi^2} - \sqrt{2n' - 1}$ may be used as a normal deviate with unit variance, where n' is the number of degrees of freedom.

Reproduced from *Statistical Methods for Research Workers*, 6th ed., with the permission of the author, R. A. Fisher, and his publisher, Oliver and Boyd, Edinburgh.

TABLE VIII

F



5% (Roman Type) and 1% (Bold-Face Type) Points for the Distribution of F

Degrees of freedom for lesser mean square	Degrees of freedom for greater mean square: 1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞
1	161	200	216	225	230	234	237	239	241	242	243	244	245	246	248	249	250	251	252	253	253	254	254	254
	4052	**4999**	**5403**	**5625**	**5764**	**5859**	**5928**	**5981**	**6022**	**6056**	**6082**	**6106**	**6142**	**6169**	**6208**	**6234**	**6258**	**6286**	**6302**	**6323**	**6334**	**6352**	**6361**	**6366**
2	18.51	19.00	19.16	19.25	19.30	19.33	19.36	19.37	19.38	19.39	19.40	19.41	19.42	19.43	19.44	19.45	19.46	19.47	19.47	19.48	19.49	19.49	19.50	19.50
	98.49	**99.01**	**99.17**	**99.25**	**99.30**	**99.33**	**99.34**	**99.36**	**99.38**	**99.40**	**99.41**	**99.42**	**99.43**	**99.44**	**99.45**	**99.46**	**99.47**	**99.48**	**99.48**	**99.49**	**99.49**	**99.49**	**99.50**	**99.50**
3	10.13	9.55	9.28	9.12	9.01	8.94	8.88	8.84	8.81	8.78	8.76	8.74	8.71	8.69	8.66	8.64	8.62	8.60	8.58	8.57	8.56	8.54	8.54	8.53
	34.12	**30.81**	**29.46**	**28.71**	**28.24**	**27.91**	**27.67**	**27.49**	**27.34**	**27.23**	**27.13**	**27.05**	**26.92**	**26.83**	**26.69**	**26.60**	**26.50**	**26.41**	**26.30**	**26.27**	**26.23**	**26.18**	**26.14**	**26.12**
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.93	5.91	5.87	5.84	5.80	5.77	5.74	5.71	5.70	5.68	5.66	5.65	5.64	5.63
	21.20	**18.00**	**16.69**	**15.98**	**15.52**	**15.21**	**14.98**	**14.80**	**14.66**	**14.54**	**14.45**	**14.37**	**14.24**	**14.15**	**14.02**	**13.93**	**13.83**	**13.74**	**13.69**	**13.61**	**13.57**	**13.52**	**13.48**	**13.46**
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.78	4.74	4.70	4.68	4.64	4.60	4.56	4.53	4.50	4.46	4.44	4.42	4.40	4.38	4.37	4.36
	16.26	**13.27**	**12.06**	**11.39**	**10.97**	**10.67**	**10.45**	**10.27**	**10.15**	**10.05**	**9.96**	**9.89**	**9.77**	**9.68**	**9.55**	**9.47**	**9.38**	**9.29**	**9.24**	**9.17**	**9.13**	**9.07**	**9.04**	**9.02**
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.96	3.92	3.87	3.84	3.81	3.77	3.75	3.72	3.71	3.69	3.68	3.67
	13.74	**10.92**	**9.78**	**9.15**	**8.75**	**8.47**	**8.26**	**8.10**	**7.98**	**7.87**	**7.79**	**7.72**	**7.60**	**7.52**	**7.39**	**7.31**	**7.23**	**7.14**	**7.09**	**7.02**	**6.99**	**6.94**	**6.90**	**6.88**
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.63	3.60	3.57	3.52	3.49	3.44	3.41	3.38	3.34	3.32	3.29	3.28	3.25	3.24	3.23
	12.25	**9.55**	**8.45**	**7.85**	**7.46**	**7.19**	**7.00**	**6.84**	**6.71**	**6.62**	**6.54**	**6.47**	**6.35**	**6.27**	**6.15**	**6.07**	**5.98**	**5.90**	**5.85**	**5.78**	**5.75**	**5.70**	**5.67**	**5.65**
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.34	3.31	3.28	3.23	3.20	3.15	3.12	3.08	3.05	3.03	3.00	2.98	2.96	2.94	2.93
	11.26	**8.65**	**7.59**	**7.01**	**6.63**	**6.37**	**6.19**	**6.03**	**5.91**	**5.82**	**5.74**	**5.67**	**5.56**	**5.48**	**5.36**	**5.28**	**5.20**	**5.11**	**5.06**	**5.00**	**4.96**	**4.91**	**4.88**	**4.86**
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.13	3.10	3.07	3.02	2.98	2.93	2.90	2.86	2.82	2.80	2.77	2.76	2.73	2.72	2.71
	10.56	**8.02**	**6.99**	**6.42**	**6.06**	**5.80**	**5.62**	**5.47**	**5.35**	**5.26**	**5.18**	**5.11**	**5.00**	**4.92**	**4.80**	**4.73**	**4.64**	**4.56**	**4.51**	**4.45**	**4.41**	**4.36**	**4.33**	**4.31**

10	4.96 **10.04**	4.10 **7.56**	3.71 **6.55**	3.48 **5.99**	3.33 **5.64**	3.22 **5.39**	3.14 **5.21**	3.07 **5.06**	3.02 **4.95**	2.97 **4.85**	2.94 **4.78**	2.91 **4.71**	2.86 **4.60**	2.82 **4.52**	2.77 **4.41**	2.74 **4.33**	2.70 **4.25**	2.67 **4.17**	2.64 **4.12**	2.61 **4.05**	2.59 **4.01**	2.56 **3.96**	2.55 **3.93**	2.54 **3.91**
11	4.84 **9.65**	3.98 **7.20**	3.59 **6.22**	3.36 **5.67**	3.20 **5.32**	3.09 **5.07**	3.01 **4.88**	2.95 **4.74**	2.90 **4.63**	2.86 **4.54**	2.82 **4.46**	2.79 **4.40**	2.74 **4.29**	2.70 **4.21**	2.65 **4.10**	2.61 **4.02**	2.57 **3.94**	2.53 **3.86**	2.50 **3.80**	2.47 **3.74**	2.45 **3.70**	2.42 **3.66**	2.41 **3.62**	2.40 **3.60**
12	4.75 **9.33**	3.88 **6.93**	3.49 **5.95**	3.26 **5.41**	3.11 **5.06**	3.00 **4.82**	2.92 **4.65**	2.85 **4.50**	2.80 **4.39**	2.76 **4.30**	2.72 **4.22**	2.69 **4.16**	2.64 **4.05**	2.60 **3.98**	2.54 **3.86**	2.50 **3.78**	2.46 **3.70**	2.42 **3.61**	2.40 **3.56**	2.36 **3.49**	2.35 **3.46**	2.32 **3.41**	2.31 **3.38**	2.30 **3.36**
13	4.67 **9.07**	3.80 **6.70**	3.41 **5.74**	3.18 **5.20**	3.02 **4.86**	2.92 **4.62**	2.84 **4.44**	2.77 **4.30**	2.72 **4.19**	2.67 **4.10**	2.63 **4.02**	2.60 **3.96**	2.55 **3.85**	2.51 **3.78**	2.46 **3.67**	2.42 **3.59**	2.38 **3.51**	2.34 **3.42**	2.32 **3.37**	2.28 **3.30**	2.26 **3.27**	2.24 **3.21**	2.22 **3.18**	2.21 **3.16**
14	4.60 **8.86**	3.74 **6.51**	3.34 **5.56**	3.11 **5.03**	2.96 **4.69**	2.85 **4.46**	2.77 **4.28**	2.70 **4.14**	2.65 **4.03**	2.60 **3.94**	2.56 **3.86**	2.53 **3.80**	2.48 **3.70**	2.44 **3.62**	2.39 **3.51**	2.35 **3.43**	2.31 **3.34**	2.27 **3.26**	2.24 **3.21**	2.21 **3.14**	2.19 **3.11**	2.16 **3.06**	2.14 **3.02**	2.13 **3.00**
15	4.54 **8.68**	3.68 **6.36**	3.29 **5.42**	3.06 **4.89**	2.90 **4.56**	2.79 **4.32**	2.70 **4.14**	2.64 **4.00**	2.59 **3.89**	2.55 **3.80**	2.51 **3.73**	2.48 **3.67**	2.43 **3.56**	2.39 **3.48**	2.33 **3.36**	2.29 **3.29**	2.25 **3.20**	2.21 **3.12**	2.18 **3.07**	2.15 **3.00**	2.12 **2.97**	2.10 **2.92**	2.08 **2.89**	2.07 **2.87**
16	4.49 **8.53**	3.63 **6.23**	3.24 **5.29**	3.01 **4.77**	2.85 **4.44**	2.74 **4.20**	2.66 **4.03**	2.59 **3.89**	2.54 **3.78**	2.49 **3.69**	2.45 **3.61**	2.42 **3.55**	2.37 **3.45**	2.33 **3.37**	2.28 **3.25**	2.24 **3.18**	2.20 **3.10**	2.16 **3.01**	2.13 **2.96**	2.09 **2.89**	2.07 **2.86**	2.04 **2.80**	2.02 **2.77**	2.01 **2.75**
17	4.45 **8.40**	3.59 **6.11**	3.20 **5.18**	2.96 **4.67**	2.81 **4.34**	2.70 **4.10**	2.62 **3.93**	2.55 **3.79**	2.50 **3.68**	2.45 **3.59**	2.41 **3.52**	2.38 **3.45**	2.33 **3.35**	2.29 **3.27**	2.23 **3.16**	2.19 **3.08**	2.15 **3.00**	2.11 **2.92**	2.08 **2.86**	2.04 **2.79**	2.02 **2.76**	1.99 **2.70**	1.97 **2.67**	1.96 **2.65**
18	4.41 **8.28**	3.55 **6.01**	3.16 **5.09**	2.93 **4.58**	2.77 **4.25**	2.66 **4.01**	2.58 **3.85**	2.51 **3.71**	2.46 **3.60**	2.41 **3.51**	2.37 **3.44**	2.34 **3.37**	2.29 **3.27**	2.25 **3.19**	2.19 **3.07**	2.15 **3.00**	2.11 **2.91**	2.07 **2.83**	2.04 **2.78**	2.00 **2.71**	1.98 **2.68**	1.95 **2.62**	1.93 **2.59**	1.92 **2.57**
19	4.38 **8.18**	3.52 **5.93**	3.13 **5.01**	2.90 **4.50**	2.74 **4.17**	2.63 **3.94**	2.55 **3.77**	2.48 **3.63**	2.43 **3.52**	2.38 **3.43**	2.34 **3.36**	2.31 **3.30**	2.26 **3.19**	2.21 **3.12**	2.15 **3.00**	2.11 **2.92**	2.07 **2.84**	2.02 **2.76**	2.00 **2.70**	1.96 **2.63**	1.94 **2.60**	1.91 **2.54**	1.90 **2.51**	1.88 **2.49**
20	4.35 **8.10**	3.49 **5.85**	3.10 **4.94**	2.87 **4.43**	2.71 **4.10**	2.60 **3.87**	2.52 **3.71**	2.45 **3.56**	2.40 **3.45**	2.35 **3.37**	2.31 **3.30**	2.28 **3.23**	2.23 **3.13**	2.18 **3.05**	2.12 **2.94**	2.08 **2.86**	2.04 **2.77**	1.99 **2.69**	1.96 **2.63**	1.92 **2.56**	1.90 **2.53**	1.87 **2.47**	1.85 **2.44**	1.84 **2.42**
21	4.32 **8.02**	3.47 **5.78**	3.07 **4.87**	2.84 **4.37**	2.68 **4.04**	2.57 **3.81**	2.49 **3.65**	2.42 **3.51**	2.37 **3.40**	2.32 **3.31**	2.28 **3.24**	2.25 **3.17**	2.20 **3.07**	2.15 **2.99**	2.09 **2.88**	2.05 **2.80**	2.00 **2.72**	1.96 **2.63**	1.93 **2.58**	1.89 **2.51**	1.87 **2.47**	1.84 **2.42**	1.82 **2.38**	1.81 **2.36**
22	4.30 **7.94**	3.44 **5.72**	3.05 **4.82**	2.82 **4.31**	2.66 **3.99**	2.55 **3.76**	2.47 **3.59**	2.40 **3.45**	2.35 **3.35**	2.30 **3.26**	2.26 **3.18**	2.23 **3.12**	2.18 **3.02**	2.13 **2.94**	2.07 **2.83**	2.03 **2.75**	1.98 **2.67**	1.93 **2.58**	1.91 **2.53**	1.87 **2.46**	1.84 **2.42**	1.81 **2.37**	1.80 **2.33**	1.78 **2.31**
23	4.28 **7.88**	3.42 **5.66**	3.03 **4.76**	2.80 **4.26**	2.64 **3.94**	2.53 **3.71**	2.45 **3.54**	2.38 **3.41**	2.32 **3.30**	2.28 **3.21**	2.24 **3.14**	2.20 **3.07**	2.14 **2.97**	2.10 **2.89**	2.04 **2.78**	2.00 **2.70**	1.96 **2.62**	1.91 **2.53**	1.88 **2.48**	1.84 **2.41**	1.82 **2.37**	1.79 **2.32**	1.77 **2.28**	1.76 **2.26**
24	4.26 **7.82**	3.40 **5.61**	3.01 **4.72**	2.78 **4.22**	2.62 **3.90**	2.51 **3.67**	2.43 **3.50**	2.36 **3.36**	2.30 **3.25**	2.26 **3.17**	2.22 **3.09**	2.18 **3.03**	2.13 **2.93**	2.09 **2.85**	2.02 **2.74**	1.98 **2.66**	1.94 **2.58**	1.89 **2.49**	1.86 **2.44**	1.82 **2.36**	1.80 **2.33**	1.76 **2.27**	1.74 **2.23**	1.73 **2.21**
25	4.24 **7.77**	3.38 **5.57**	2.99 **4.68**	2.76 **4.18**	2.60 **3.86**	2.49 **3.63**	2.41 **3.46**	2.34 **3.32**	2.28 **3.21**	2.24 **3.13**	2.20 **3.05**	2.16 **2.99**	2.11 **2.89**	2.06 **2.81**	2.00 **2.70**	1.96 **2.62**	1.92 **2.54**	1.87 **2.45**	1.84 **2.40**	1.80 **2.32**	1.77 **2.29**	1.74 **2.23**	1.72 **2.19**	1.71 **2.17**
26	4.22 **7.72**	3.37 **5.53**	2.89 **4.64**	2.74 **4.14**	2.59 **3.82**	2.47 **3.59**	2.39 **3.42**	2.32 **3.29**	2.27 **3.17**	2.22 **3.09**	2.18 **3.02**	2.15 **2.96**	2.10 **2.86**	2.05 **2.77**	1.99 **2.66**	1.95 **2.58**	1.90 **2.50**	1.85 **2.41**	1.82 **2.36**	1.78 **2.28**	1.76 **2.25**	1.72 **2.19**	1.70 **2.15**	1.69 **2.13**

TABLE VIII — *Continued*

F

5% (Roman Type) and 1% (Bold-Face Type) Points for the Distribution of *F*

Degrees of freedom for lesser mean square	Degrees of freedom for greater mean square: 1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.30	2.25	2.20	2.16	2.13	2.08	2.03	1.97	1.93	1.88	1.84	1.80	1.76	1.74	1.71	1.68	1.67
	7.68	**5.49**	**4.60**	**4.11**	**3.79**	**3.56**	**3.39**	**3.26**	**3.14**	**3.06**	**2.98**	**2.93**	**2.83**	**2.74**	**2.63**	**2.55**	**2.47**	**2.38**	**2.33**	**2.25**	**2.21**	**2.16**	**2.12**	**2.10**
28	4.20	3.34	2.95	2.71	2.56	2.44	2.36	2.29	3.24	2.19	2.15	2.12	2.06	2.02	1.96	1.91	1.87	1.81	1.78	1.75	1.72	1.69	1.67	1.65
	7.64	**5.45**	**4.57**	**4.07**	**3.76**	**3.53**	**3.36**	**3.23**	**3.11**	**3.03**	**2.95**	**2.90**	**2.80**	**2.71**	**2.60**	**2.52**	**2.44**	**2.35**	**2.30**	**2.22**	**2.18**	**2.13**	**2.09**	**2.06**
29	4.18	3.33	2.93	2.70	2.54	2.43	2.35	2.28	2.22	2.18	2.14	2.10	2.05	2.00	1.94	1.90	1.85	1.80	1.77	1.73	1.71	1.68	1.65	1.64
	7.60	**5.52**	**4.54**	**4.04**	**3.73**	**3.50**	**3.33**	**3.20**	**3.08**	**3.00**	**2.92**	**2.87**	**2.77**	**2.68**	**2.57**	**2.49**	**2.41**	**2.32**	**2.27**	**2.19**	**2.15**	**2.10**	**2.06**	**2.03**
30	4.17	3.32	2.92	2.69	2.53	2.42	2.34	2.27	2.21	2.16	2.12	2.09	2.04	1.99	1.93	1.89	1.84	1.79	1.76	1.72	1.69	1.66	1.64	1.62
	7.56	**5.39**	**4.51**	**4.02**	**3.70**	**3.47**	**3.30**	**3.17**	**3.06**	**2.98**	**2.90**	**2.84**	**2.74**	**2.66**	**2.55**	**2.47**	**2.38**	**2.29**	**2.24**	**2.16**	**2.13**	**2.07**	**2.03**	**2.01**
32	4.15	3.30	2.90	2.67	2.51	2.40	2.32	2.25	2.19	2.14	2.10	2.07	2.02	1.97	1.91	1.86	1.82	1.76	1.74	1.69	1.67	1.64	1.61	1.59
	7.50	**5.34**	**4.46**	**3.97**	**3.66**	**3.42**	**3.25**	**3.12**	**3.01**	**2.94**	**2.86**	**2.80**	**2.70**	**2.62**	**2.51**	**2.42**	**2.34**	**2.25**	**2.20**	**2.12**	**2.08**	**2.02**	**1.98**	**1.96**
34	4.13	3.28	2.88	2.65	2.49	2.38	2.30	2.23	2.17	2.12	2.08	2.05	2.00	1.95	1.89	1.84	1.80	1.74	1.71	1.67	1.64	1.61	1.59	1.57
	7.44	**5.29**	**4.42**	**3.93**	**3.61**	**3.38**	**3.21**	**3.08**	**2.97**	**2.89**	**2.82**	**2.76**	**2.66**	**2.58**	**2.47**	**2.38**	**2.30**	**2.21**	**2.15**	**2.08**	**2.04**	**1.98**	**1.94**	**1.91**
36	4.11	3.26	2.86	2.63	2.48	2.36	2.28	2.21	2.15	2.10	2.06	2.03	1.89	1.93	1.87	1.82	1.78	1.72	1.69	1.65	1.62	1.59	1.56	1.55
	7.39	**5.25**	**4.38**	**3.89**	**3.58**	**3.35**	**3.18**	**3.04**	**2.94**	**2.86**	**2.78**	**2.72**	**2.62**	**2.54**	**2.43**	**2.35**	**2.26**	**2.17**	**2.12**	**2.04**	**2.00**	**1.94**	**1.90**	**1.87**
38	4.10	3.25	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	2.05	2.02	1.96	1.92	1.85	1.80	1.76	1.71	1.67	1.63	1.60	1.57	1.54	1.53
	7.35	**5.21**	**4.34**	**3.86**	**3.54**	**3.32**	**3.15**	**3.02**	**2.91**	**2.82**	**2.75**	**2.69**	**2.59**	**2.51**	**2.40**	**2.32**	**2.22**	**2.14**	**2.08**	**2.00**	**1.97**	**1.90**	**1.86**	**1.84**
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.07	2.04	2.00	1.95	1.90	1.84	1.79	1.74	1.69	1.66	1.61	1.59	1.55	1.53	1.51
	7.31	**5.18**	**4.31**	**3.83**	**3.51**	**3.29**	**3.12**	**2.99**	**2.88**	**2.80**	**2.73**	**2.66**	**2.56**	**2.49**	**2.37**	**2.29**	**2.20**	**2.11**	**2.05**	**1.97**	**1.94**	**1.88**	**1.84**	**1.81**
42	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11	2.06	2.02	1.99	1.94	1.89	1.82	1.78	1.73	1.68	1.64	1.60	1.57	1.54	1.51	1.49
	7.27	**5.15**	**4.29**	**3.80**	**3.49**	**3.26**	**3.10**	**2.96**	**2.86**	**2.77**	**2.70**	**2.64**	**2.54**	**2.46**	**2.35**	**2.26**	**2.17**	**2.08**	**2.02**	**1.94**	**1.91**	**1.85**	**1.80**	**1.78**
44	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10	2.05	2.01	1.98	1.92	1.88	1.81	1.76	1.72	1.66	1.63	1.58	1.56	1.52	1.50	1.48
	7.24	**5.12**	**4.26**	**3.78**	**3.46**	**3.24**	**3.07**	**2.94**	**2.84**	**2.75**	**2.68**	**2.62**	**2.52**	**2.44**	**2.32**	**2.24**	**2.15**	**2.06**	**2.00**	**1.92**	**1.88**	**1.82**	**1.78**	**1.75**
46	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.14	2.09	2.04	2.00	1.97	1.91	1.87	1.80	1.75	1.71	1.65	1.62	1.57	1.54	1.51	1.48	1.46
	7.21	**5.10**	**4.24**	**3.76**	**3.44**	**3.22**	**3.05**	**2.92**	**2.82**	**2.73**	**2.66**	**2.60**	**2.50**	**2.42**	**2.30**	**2.22**	**2.13**	**2.04**	**1.98**	**1.90**	**1.86**	**1.80**	**1.76**	**1.72**
48	4.04	3.19	2.80	2.56	2.41	2.30	2.21	2.14	2.08	2.03	1.99	1.96	1.90	1.86	1.79	1.74	1.70	1.64	1.61	1.56	1.53	1.50	1.47	1.45
	7.19	**5.08**	**4.22**	**3.74**	**3.42**	**3.20**	**3.04**	**2.90**	**2.80**	**2.71**	**2.64**	**2.58**	**2.48**	**2.40**	**2.28**	**2.20**	**2.11**	**2.02**	**1.96**	**1.88**	**1.84**	**1.78**	**1.73**	**1.70**

INDEX